A study of the reliability and validity of tests of divergent thinking.

Richards, P. N.

Award date:
1977

Awarding institution:
University of Bath

Link to publication

A STUDY OF THE RELIABILITY AND VALIDITY OF TESTS OF DIVERGENT THINKING

submitted by P.N. Richards

for the degree of Ph.D.

of the University of Bath.

1977

P. Richards

ACKNOWLEDGEMENTS

I should like to acknowledge the advice and encouragement of my supervisor, Professor K. Austwick; the assistance of Keith Pledger and Anthony King in carrying out the Board Game and inter-scorer reliability, respectively; and the cooperation and hospitality of the headmasters and staffs of the schools concerned.

I also wish to express particular thanks to Hazel Gott for all her work in carrying out the typing and for the able and painstaking way in which she has transformed my draft into the present copy.

ABSTRACT

A great deal of psychological and educational practice over the last twenty years has been based on the belief that creative abilities are present to some extent in all children, and tests of divergent thinking (D.T.) have been widely used as criteria of 'creativity' in the research literature.

The first part of this study considers the theoretical background to the above statement, and expresses some doubts about the reliability and validity of D.T. tests. The identification of D.T. tests as tests of 'creativity' was considered unwarranted though some theoretical association was claimed in this respect, and a review of previous researches lent some support to this view.

Considering the problems involved in scoring D.T. tests, the varied information on reliability and construct validity, and the limited reports of long-term stability and criterion-related validity, an attempt was made to investigate each of these aspects, with three D.T. tests, 'Circles', 'Uses', and 'Consequences'. The main investigation involved 161 eleven-year-old children, and the long-term stability results were obtained from a follow-up study after nearly five years, of 139 pupils of 15 and 16 years of age.

The findings gave some positive support to each aspect investigated, though limitations and variations between different tests and between the sexes were also observed. In general the D.T. tests emerged, after considerable scrutiny, as reasonably reliable measures of an intellectual ability which was relatively independent of intelligence, showed development and stability of relative ranking over time, and was positively related to creative behaviour.

Though the identification of D.T. tests as 'creativity' tests was rejected, both theoretically and experimentally, a limited amount of association between D.T. and creative behaviour was demonstrated, and the evidence should provide some basic support for teachers who regard the fostering of D.T. abilities as a worthwhile aim.

CONTENTS

Acknowledgements
Abstract
Contents

Page

CHAPTER ONE

THE RELIABILITY AND VALIDITY OF DIVERGENT THINKING TESTS: PROBLEMS AND PERSPECTIVES 1

1.1 Introduction 1
1.2 Validity 6
1.21 Concurrent and Predictive Validity 10
1.22 Construct Validity 16
1.3 Reliability 23
1.31 Test-retest Reliability 25
1.32 Alternate-form Reliability 27
1.33 Internal Consistency 27
1.34 Discussion 30
1.35 Long-term Stability 38
1.36 Scorer Reliability 39
1.4 Summary 42

CHAPTER TWO

CREATIVITY, DIVERGENT THINKING AND CLASSROOM LEARNING 44

2.1 Creativity and the Development of Creative Abilities in the Classroom 44
2.11 Defining Creativity 46
2.12 Educational Implications 55
2.2 Divergent Thinking Tests - not 'Creativity' Tests 59

CHAPTER THREE

THEORETICAL APPROACHES TO CREATIVE THINKING 67

3.1 Divergent Thinking and Guilford's Structure-of-Intellect Theory 68
3.11 The Category of Divergent Thinking 71
3.12 Discussion 76
3.2 Creative Thinking and Problem Solving 80
3.3 Unconscious Drives and Motives in Creative Thinking 84
3.4 Association and Creative Production 89
3.5 Gestalt Approaches to Creative Thinking 95
3.6 Styles of Thinking 100

Page

3.7 Psychometric Approaches to Creative Thinking .. 107

3.8 Conclusion 118

CHAPTER FOUR

REVIEW OF INVESTIGATIONS INTO THE VALIDITY AND RELIABILITY OF DIVERGENT THINKING TESTS 121

4.1 Introduction 121

4.2 Empirical Evidence for the Validity of Divergent Thinking Tests based on the Claims of Torrance (1974) 122

4.3 Further Studies of the Validity of Divergent Tests 132

4.4 Studies of the Relationship between Divergent Thinking and Intelligence 143

4.5 Evidence for the Test-retest Reliability of Divergent Thinking Tests 147

4.6 Studies of Validity and Reliability by Bennett (1973), Dewing (1970) and Vernon (1972) 149

4.7 Summary 159

CHAPTER FIVE

EXPERIMENTAL PROCEDURES 161

5.1 Introduction 161

5.2 The Main 11-year-old Sample 161

5.3 Testing Material 164

5.31 Divergent Thinking Tests 164

5.311 Circles Test 164

5.312 Uses Test 165

5.313 Consequences Test 166

5.314 Representative Nature of the Tests .. 167

5.32 Criteria of Creative Interests and Behaviour 170

5.321 Teacher Ratings of Creative Behaviour 170

5.322 Creative Leisure Activities Checklist 171

5.323 Interests 171

5.324 Creative Motivation Scale 173

5.325 The Board Game 173

5.3251 Introduction 173

5.3252 Description and Administration 175

5.3253 Development and Assessment of Performance 178

5.3254 Categories of Performance .. 181

Page

5.33 Convergent Thinking 183

5.331 Intelligence (VRQ) 183

5.332 Academic Performance in English and Mathematics 184

5.34 Personality and Attitudes 185

5.341 Junior Eysenck Personality Inventory 185

5.342 Self Concept 185

5.343 Teacher Ratings of Children's Personality and Pleasurability 186

5.344 Attitudes to School and School Work .. 186

5.4 Organisation 190

5.41 General Organisation 190

5.42 Testing Sessions 191

5.43 Preparatory Lesson and the Board Game 192

CHAPTER SIX

THE SCORING OF DIVERGENT THINKING TESTS 195

6.1 Introduction 195

6.2 The Reporting of Scoring Procedures in Divergent Thinking Research 198

6.3 Review and Practical Procedures for Scoring for Fluency, Flexibility and Originality 201

6.31 Fluency 201

6.311 Relevance of Responses 205

6.32 Flexibility 210

6.321 Practical Details of Flexibility Scoring 215

6.33 Originality 219

6.331 Use of a Sample of Scripts for Listing Uncommon Responses 224

6.332 Summary 227

6.333 Practical Details of Originality Scoring 228

6.4 A Note on the Time Involved in Scoring Divergent Tests 234

6.5 Synopsis of Scoring Details for the Divergent Thinking Tests 238

6.51 Circles Test 238

6.52 Squares Test 240

6.53 Uses Test 241

6.54 Consequences Test 242

6.55 Selection of the Distribution Sample for Assessing Uncommon Responses 243

6.56 Examples of Common and Uncommon Responses to Each of the Divergent Tests 243

Page

CHAPTER SEVEN

INTERSCORER RELIABILITY 246

7.1 Introduction 246

7.2 Scorer Reliability in the Present Study 253

7.21 Selection of the Sample 253

7.22 Procedure 254

7.23 Results and Discussion 255

7.3 Conclusion 263

CHAPTER EIGHT

TEST-RETEST RELIABILITY OF TESTS OF DIVERGENT THINKING 266

8.1 Introduction 266

8.2 Results and Discussion 268

8.21 Level of Reliability 268

8.22 Comparison with other Studies 270

8.23 Levels of Performance from Test to Retest .. 275

8.24 Reliability Coefficients and Distribution of Variance 279

8.3 Summary 283

CHAPTER NINE

LONG-TERM STABILITY OF DIVERGENT THINKING 285

9.1 Background 285

9.2 Tests and Procedures 290

9.21 Divergent Tests 290

9.22 Intelligence Tests 291

9.23 Administration 291

9.3 Results 293

9.31 Introduction 293

9.32 Sex Differences in Divergent Thinking - 1969 and 1974 294

9.33 Long-term Test-retest Reliabilities 1969-1974 296

9.331 296

9.332 Comparison with Previous Researches .. 298

9.34 Levels of Performance on Divergent Tests 1969-1974 302

9.35 Correction for Attenuation 304

9.36 Relationship with Intelligence 308

9.37 Intercorrelations between Divergent Thinking Tests 1969 and 1974 312

9.4 Summary 317

Page

CHAPTER TEN

RESULTS OF THE INVESTIGATION OF THE CONCURRENT VALIDITY OF TESTS OF DIVERGENT THINKING 319

10.1 Introduction 319

10.2 Results and Discussion 321

10.21 Intercorrelations between the Divergent Tests and between the Creative Criteria .. 321

10.22 Basic Validity Coefficients 325

10.23 An Appraisal of the Validity Coefficients for each Divergent Thinking Measure in Conjunction with Sex Differences and the Effect of Intelligence 335

10.231 Validity Coefficients with Teacher Ratings of Children's Creative Behaviour 335

10.232 Validity Coefficients with Torrance's Leisure Interests Checklist 340

10.233 Validity Coefficients with Golann's Creative Motivation Scale 344

10.234 Validity Coefficients with Imaginative Interests 348

10.235 Validity Coefficients with Logical/ Analytic Interests 351

10.236 Concurrent Validity of Divergent Thinking Tests in Relation to the Board Game 354

10.24 Categorical Analysis of the Validity of the Divergent Tests in Relation to the Board Game 358

10.25 Relative Performance of Convergers and Divergers on the Board Game 367

10.26 An Illustration of the Potential of Divergent Thinking Tests and I.Q. in Forming a Multiple Validity Coefficient 371

10.3 Summary and Conclusions 374

10.31 General Evidence for Concurrent Validity .. 374

10.32 The Effect of Intelligence on the Validity Coefficients of the Divergent Tests 375

10.33 Sex Differences and Validity Coefficients with Individual Criteria 377

10.34 The Board Game 380

10.35 Individual Variations in the Validity of the Divergent Tests 381

10.36 General Appraisal of the Evidence Presented 385

Page

CHAPTER ELEVEN

RESULTS OF THE FACTORIAL INVESTIGATION OF CONSTRUCT VALIDITY 390

11.1 Introduction 390

11.2 Results and Discussion 398

11.21 Identification of the Factors 401

11.22 Discussion 402

11.221 Convergent Thinking and Divergent Thinking 402

11.222 Attitudes and Divergent Thinking .. 408

11.223 Creative Activities and Divergent Thinking 413

11.224 Divergent Thinking and Personality Variables 417

11.3 Summary and Conclusions 422

CHAPTER TWELVE

OVERVIEW AND CONCLUSION 427

REFERENCES 442

APPENDIX 459

1. Data Analysis

Long-term Stability:
Means, Standard Deviations and Intercorrelations 460

Main 11-year-old Analysis:
Means and Standard Deviations, Boys, Girls and Whole Population 461

Intercorrelations between all Variables:

(i) Girls (n = 84) 462

(ii) Boys (n = 77) 463

(iii) Whole population (n = 161) 464

Factor Analysis:

(i) Girls 465

(ii) Boys 466

Interscorer Reliability:
Marks Awarded by Scorer 1 and Scorer 2 .. 467

2. Samples of Materials Used in Scoring the Divergent Tests

Flexibility Categories for the Circles Test and the Squares Test 470

Frequency Distribution of Responses to the Circles Test 472

Page

Frequency Distribution of Responses to the Squares Test 480

Frequency Distribution of Responses to the Uses Test 484

Frequency Distribution of Responses to the Consequences Test 487

3. Board Game

Photographs of each Category of Response 490

Pupil's 'report' on the Preliminary Lesson 493

Record Sheet for Observing Children's Performance 494

4. Test Materials

(i) Divergent Thinking Booklet (as used for the 15-year-old group) & 'Squares' 495

(ii) Attitudes: 'What I think about School' 501

(iii) Self Concept: 'What I am Like' 507

(iv) Teacher Ratings. (a) 'Children's School Work and Personality' 508

(b) 'Indicators of Creative Behaviour' 510

(v) Creative Motivation Scale: 'Things I Would Like to do Most' 512

(vi) Creative Leisure Interests: 'Things You have done in your Spare Time' .. 515

(vii) Interests (Imaginative and Logical/Analytic) .. 519

CHAPTER ONE

THE RELIABILITY AND VALIDITY OF DIVERGENT THINKING TESTS: PROBLEMS AND PERSPECTIVES

1.1 INTRODUCTION

Based largely on the rationale presented by Guilford (1950) and encouraged by the relevance to the classroom of many of the creativity writings of the early 1960's (Torrance, 1962, 1963; Getzels & Jackson, 1962; Hudson, 1966), tests of divergent thinking have been regarded, on both sides of the Atlantic, as potentially valuable instruments in a climate of educational interest in creative ability.

With the growing demands of science and technology for people with the ability to think openly and creatively, and the pressures on education to encourage and develop such talents, it is not surprising that Guilford's plea for a renewal of interest in creativity, coupled with his suggestions for identifying and measuring creative abilities, should have become the springboard for the vast increase in educational and psychological research into creative abilities that has taken place since his now famous Presidential address on 'Creativity' to the American Psychological Association (Guilford, 1950).

Guilford first established the word 'divergent' to describe a particular area of thinking abilities in his 'structure of intellect model' (Guilford, 1956). In this he formalises views he had expressed earlier that a study of intellectual abilities needs a multidimensional approach which should include components of creative talent such as originality, fluency of ideas, and flexibility, including the ability to change direction in one's thinking, to depart from well-worn paths, and to revise the apparent restrictions of rigid rules and methods. These abilities he places firmly in a category of divergent thinking.

Similar abilities have figured prominently in accounts of creative expression and invention over the centuries, though the sources of inspiration and originality were often regarded as being beyond the comprehension of ordinary men, and to be characteristic of a race of men apart, whose creativity manifested itself in some recognisable product of the highest order. Since Galton (1869), in his investigations into hereditary genius, suggested that such abilities are a matter of degree rather than kind, more attention has been given to the measurement and development of exceptional talent.

Before the stimulus of Guilford's (1950) paper however most investigations of creative talent this century concentrated on high intelligence as the supreme index of creative ability, and while, as even Getzels and Jackson (1962) acknowledge "IQ ...... remains one of the best predictors of academic achievement we have", there has been growing concern that too great an emphasis on the type of logical activity associated with intelligence tests may discourage spontaneous and imaginative thinking of a kind which might contribute to the development of future original creation.

Guilford (1956) suggests that the concept of intelligence should be widened to cover abilities of both types, and in his structure of intellect model includes a wide range of abilities in five categories, cognition, memory, divergent production, convergent production and evaluation. He claims, however, that "it is in the divergent thinking category that we find the abilities that are most significant in creative thinking and invention" (Guilford, 1958). Such claims have led to a widespread use of divergent thinking tests in investigations of creative abilities to the extent that they have tended to become synonymous with tests of creativity in the research literature. The creativity tests of Torrance alone (Torrance, 1962, 1974) are among the testing instruments most frequently cited by publications in the latest, 7th, Mental Measurements Yearbook (Buros, 1972). The theoretical justification for such tests is appealing, but as yet there is little real evidence to support it, and the tests themselves have been subject to far less investigation than one might assume considering their predominance.

It is hoped that the present study will provide some worthwhile evidence regarding both their validity and reliability. It must be pointed out however that the whole psychological study of creativity covers a much wider range than studies based simply on the use of divergent thinking tests. Reviewing the field, Golann (1963) emphasises the diversity of interests, motives and approaches characteristic of the many investigations of creative ability. He suggests that four main emphases are apparent: an emphasis on creative productivity in technological and industrial settings; the creative process and its associated mental activity; measurement, focussing on the work of Guilford and Torrance; and personality studies including psychodynamic approaches and personality characteristics. A further emphasis on children's thinking and school learning has appeared to build up since Golann's review.

The emphasis in the present study is on the role of divergent thinking as indices of a mental activity related to a 'process' definition of creativity. That it is also relevant to the classroom is implicit in its consideration of children's thinking, but this link will be made specific, in the view, as suggested by Vernon (1964), that schools can do much to stimulate and foster, or else to inhibit, creative talent. It is also the view of the writer that such talent involves skills of divergent thinking and that these skills can be developed in the classroom. Some recent research by the writer and B.D. Franklin reported elsewhere (Franklin and Richards, 1977) lends some support to this view.

It is suggested that research into particular divergent thinking skills is likely in practice to be more profitable than research done under the umbrella title of 'creativity'. Different definitions, theories and means of evaluation of 'creativity' have in fact been responsible for much of what Yamamoto (1965) describes as the 'confused abundance', and others, (Klein *et al*, 1967), the 'complete chaos' of creativity research. Divergent thinking tests have contributed to the confusion by being widely used as indices and criteria of creativity when in the opinion of some researchers "there is scarcely a shred of factual support for this". (Hudson, 1966).

It is suggested by Cronbach (1970) that although the greatest amount of creativity research has accumulated around divergent thinking tests, especially those designed by Torrance (1959, 1962), "most of the tests have been announced to the world prior to any solid validity studies". More quantitatively, Nuttall (1972), claims that until recently for every ten studies using divergent tests and accepting the validity of the title 'creativity', only one attempted an empirical validation of the tests against criteria of creativity and few of these met with any great success.

While not disparaging further research into relationships between divergent tests and other abilities, both Burt (1962) and Vernon (1964) also question the tests validity and reliability, and more specifically Cropley (1967) points out that such tests have been shown to have "unsatisfactorily low reliabilities".

Brown (1976) commenting on the evidence for validity and reliability concludes that the reliability is "lower than desirable", and that "no clear pattern of validity results has emerged". Butcher (1972) referring to creativity in terms of "creativeness in real life as shown by, for example, scientific or artistic achievement" questions the assumption that high ability on divergent thinking tests is evidence of such achievement. He also questions whether various measures of divergent thinking inter-correlate sufficiently highly to justify the assumption of a general factor of divergent thinking and whether any general factor of this kind is readily distinguishable from the corresponding factor of convergent thinking.

Three issues emerge from this preliminary discussion; the appropriateness of the labelling of divergent thinking tests as tests of creativity, doubts about their validity and reliability, and the relevance of divergent thinking to the educational process in the classroom. The three issues are of course inter-related. The first two are in fact closely inter-dependant, and the third provides the practical environment in which educational psychology is put into practice in a deliberate attempt to recognise and develop childrens' abilities.

Each of these three issues will be looked at more closely, beginning in this chapter with the concepts of validity and reliability.

1.2 VALIDITY

The validity of a test is broadly defined as the capacity of a test to predict some specified behavioural measure, or set of measures other than itself (Cattell, 1964). More generally, it is expressed in terms of how well a test measures what it is supposed to measure. Procedures for determining test validity are therefore concerned with the relationships between performance on the test and some other independently observable facts about the behaviour characteristics under consideration. As the external data can be arrived at from a variety of sources, validity is not some general pervasive characteristic that a test 'has'. Rather it is an index of how well the test is likely to do for a specific purpose in a particular situation. This is particularly important when considering the variety of comments made about the validity of divergent thinking tests in relation to claims made for them as tests of 'creativity'.

Several types of validity have already been referred to indirectly in the criticisms of divergent thinking tests made above. Thus Butcher refers to both 'predictive validity' and 'construct validity' when talking of 'real creativeness in later life' and 'intercorrelations between various divergent thinking measures', respectively. Nuttall's remarks on 'empirical validation against creative criteria' were questioning 'concurrent validity', and Guilford's observation that the tests 'are constructed to reflect the sorts of abilities that are significant in creative thinking and invention' makes claims for, though do not demonstrate construct validity. A number of other descriptions of validity could also be applied to divergent tests though there would be considerable overlap with the divisions mentioned.

In an attempt to standardise the reporting of reliability and validity the American Psychological Association and the American Educational Research Association set up a joint committee in 1953 to suggest a basic set of concepts for future use in evaluating the reliability and validity of educational and psychological tests. This led to the formulation of each Association's 'Technical Recommendations' (A.P.A., 1954; A.E.R.A., 1955) and subsequently to a combined and revised version, most recently published as 'Standards for Educational and Psychological Tests' (A.P.A., 1974). The 'Standards' have resulted in the widespread classification of validity procedures into four categories of content, concurrent, predictive and construct validity. (Educational Testing Service, 1961; Anastasi, 1968; Cronbach, 1970, 1971.) As both Anastasi and Cronbach point out, concurrent and predictive validity are included in a single category of criterion-related validity in the first

(1966) and subsequent versions of the 'Standards', but their interpretations of these two aspects vary somewhat and will be discussed further later. Meanwhile the three categories of concurrent, predictive and construct validity will be used as a basis for the investigations in this study.

Some individuals however have found the above categories, defined as they are at some common denominator of meaning, to be too unsophisticated for their purpose. Cattell (1964), for example, suggests that more basic concepts are needed to match some of the increasingly subtle concepts involved in personality and motivation measurement, and Cronbach, himself a member of the A.P.A. committee, felt it necessary to clarify his views regarding their introduction of the category of construct validity. (Cronbach and Meehl, 1955, 1972.) With certain types of test some of these categories will obviously be less important than others, and additional measures may be needed to provide a useful basis from which to evaluate the tests. There will be occasions in this study when some different types of validity measurements will be informative, and these will be introduced where appropriate. On the other hand, content validity, which is associated with adequate coverage of content and objectives in the case of achievement tests is not appropriate in evaluating divergent tests.

Although it would be possible to stretch the concept of 'content validity' to a consideration of the different facets of divergent thinking sampled by any particular test, the question of the nature of the variable measured by a psychological test is answered by most writers (Brown, 1976) in terms of the evidence from studies of construct validity.

Content validity in this study is therefore subsumed under the concept of 'construct validity' to be discussed later.

Often associated with content validity is a concept of 'face validity', though Anastasi (1968) warns that this is not validity in the technical sense and should not be regarded as a substitute for objectively determined validity. It is basically a judgement process based on the appearance of the test to be measuring what it claimed to measure. Garrett (1966) notes that it is nevertheless a useful consideration in the construction of tests where there is no easily defined area of content such as for tests and rating scales for various hypothesised traits. At first sight this appears very relevant to divergent thinking tests particularly if such judgements included an appraisal of the thinking process appearing to underly the questions. Anastasi (1968) however notes that face validity is not a matter concerning technical scrutiny but a superficial assessment of the tests appropriateness by the subjects themselves, administrators and "other technically untrained observers". It therefore is more appropriate to consider the psychological appearance of the test with the theoretical foundations under construct validity.

It is worth noting however that some of the questions in divergent tests - 'how many uses can you think of for a brick', 'tell me all the ways in which a cat and a mouse are alike' - may appear silly or inappropriate, even to young people, and if so, then face validity is an important consideration which would deserve investigation. It is also worth noting on the other hand that in certain tests the true nature of the items needs to be concealed, e.g. in tests of dogmatism or personality, and the less obvious the face validity of the items the less likely are the subjects to misrepresent themselves.

The three remaining categories of concurrent, predictive and construct validity are more relevant to investigations of divergent thinking tests, and these will now be looked at more closely.

1.21 Concurrent Validity and Predictive Validity

These two types of validity have been taken together as they are both aspects of what is often termed 'criterion-related' validity. In general this indicates the effectiveness of a test in predicting an individual's behaviour in certain specified situations. For this purpose, performance on the test is checked against a 'criterion', the latter being a direct and independent measure of that which the test is designed to predict.

Prediction in this broad sense refers to prediction from the test to any criterion situation, whether the latter is obtained after some significant time interval or 'concurrently', i.e. at the same, or nearly the same time. Anastasi (1968) however suggests that it is in the former more limited sense of prediction over a time interval that the term predictive valdity is generally used. Cronbach (1970) takes a slightly different view. He suggests that "logically, predictive and concurrent validation are the same, and most writers apply the term 'predictive' to both." If however one wishes to emphasise that no time has elapsed between measures, he agrees that the study can be spoken of as a 'concurrent' validation.

Strictly speaking, and in the sense in which it is commonly used in scientific experiments, prediction is no doubt as Cronbach suggests a matter of prediction from one situation or property to another, not necessarily involving a projection into the future. Common usage however appears to favour Anastasi's interpretation. For the sake of clarity both terms will be retained in the present study with the time element as their distinguishing feature. It will still be appropriate to talk of the concurrent validity of a test in predicting something other than itself.

Questions of prediction over long periods of time will also be discussed, though in this case, as with the concept of 'stability', to be discussed later, it is acknowledged that it is the trait rather than the test itself which is being investigated. It is valuable for such information to be included in evaluations of test reliability and validity, though it is possible for misunderstandings to arise if they are interpreted strictly as coefficients of validity and reliability for the test itself. Cattell (1964) points out that some tests, especially of attitudes or motivation, are known to have quite a high degree of predictive validity or reliability over short periods of time, but to involve traits which change markedly with increasing age. To report coefficients of long term stability or predictive validity as evaluative measurements of such tests would not be a fair measure of their reliability or validity in the more restricted sense normally required in test manuals. Cattell does not suggest that long-term information is without value, but suggests that alternative labels might be more appropriate, such as 'dependability' or 'scientific utility' respectively. Such terms are not in general use however and having clarified the interpretation of long term stability and predictive validity the latter terms will be maintained in this study.

Returning to the use of the term 'concurrent validity' it is important to point out that although it is often easier to assess than predictive validity it should not be interpreted as an inferior form of test validation. It is true, as Pilliner (1968), writing on examinations, points out, that because of the time-lag and difficulties of extending validation procedures over the time required for predictive validity "test constructors often substitute 'concurrent' procedures". His conclusion that "concurrent validation for examinations is scarcely a viable procedure" should however be seen only in terms of examinations as predictors of future outcomes. A useful discussion of validity and reliability as they apply to school examinations is given by Thyne (1974). Increasingly the purposes of tests and examinations are diagnositic rather than predictive and in such cases as Anastasi (1968) points out "concurrent validity is the most appropriate type and can be justified in its own right". Provided the diagnostic groups have distinguishing features that could be revealed in some other way, the concurrent validity of the test in predicting this can be assessed. Thus, for example, it would be useful to have a test which would diagnose children's mathematical difficulties even though the same difficulties could be revealed by personal attention from the teacher, given sufficient time.

Hudson (1968) justifies the validity of divergent thinking tests in a somewhat similar way. He suggests that if a testing "differentiates between variables of psychological or social interest - if it differentiates arts specialists from scientists, let us say - then it interests us; if it does not, it does not". He therefore justifies their use in terms of "their external validity: their power to differentiate among variables other than themselves". At the stage at which Hudson is writing however his claims are still at the exploratory stage of suggesting the usefulness of the tests in discriminating between people with certain characteristics which only might prove pervasive. He is thus exploring what will shortly be termed the 'construct validity' of the tests rather than their concurrent validity. If at a later stage one of his hypotheses, for example, that divergent thinking tests can discriminate between Arts and Science bias is evaluated by further experiment, then the tests concurrent validity is being examined for that specific purpose.

A number of researchers have since examined some of Hudson's findings (e.g. Child, 1968; Mackay and Cameron, 1968) and provided some concurrent validity in relation to the characteristics of convergers and divergers. Others (e.g. Christie, 1970) have not been so successful and as Butcher (1968) and Lewis (1974) point out there is still need for some caution in interpreting Hudson's results. In particular, in the English education system Butcher suggests that the degree of specialisation might even be the cause rather than the effect of the two modes of thinking investigated by Hudson. It is worth noting however that Hudson makes no claim for the validity of divergent tests in relation to creative ability. In fact he notes that for his purposes such a consideration would be a "technical red herring" (Hudson, 1966).

The most common reference to divergent thinking tests is, nevertheless, that they are 'creativity' tests, and so common has this nomenclature become, that they are often accepted unquestioningly as having validity in predicting creative work in real life. Evidence that has been reported in support of this claim, and studies that have investigated such validity are reviewed later. Putting the question of validity in perspective however, it would be unrealistic to expect divergent thinking ability to appear as an effective single predictor of some creative criterion. Divergent thinking tests would still be valuable if they added something to the predictive value of some battery of tests over and above that covered by more conventional tests, such as I.Q. tests. In this sense it is their 'incremental' validity (Sechrest, 1963) which is important - the extent to which they raise the multiple correlation between the predictors and some criterion (Lewis, 1974). Even if, as Cronbach (1970) points out, a criterion can be predicted only with validity 0.20, the test may still make an appreciable practical contribution, and this is especially true if it is part of a battery containing independent tests measuring different factors. "Success," he writes, "is never unidimensional", and the writer's personal view is that divergent thinking abilities may well provide some of the elements needed for creative achievement.

Criteria that have been used for concurrent and predictive validity include ratings of individual's creative activity at work, school, or home, self-ratings of creative interests and activities, and academic achievement including imaginative story-writing, and various end of course results. Any test may, as Anastasi (1968) points out, "be validated against as many criteria as there are specific uses for it", but it is essential to recognise that validity in one context does not imply validity in another.

The relation between test scores and criterion measures can be expressed in a number of ways, including simple descriptions of the test scores of individuals who either perform 'well' or 'badly' on some job, course of study, or other criterion. The most common method however is the calculation of a 'validity coefficient' - the correlation between test score and criterion measure. The normal statistical requirements and limitations need to be applied to whatever method is used to calculate the coefficient.

Criteria to be used in the present study

In the present study a number of criterion measures will be selected to investigate the validity of divergent tests. Although they will not include real creative performance in later life, they will look at prediction in terms of a number of concurrent criteria. Long-term stability of the concept will also be looked at and this will be of relevance to any projection into the future of the concurrent findings.

The criteria will include ratings of the children's creative behaviour by teachers and from self-report inventories of their creative interests. In particular an experimental exercise, the 'Board Game' is to be designed, in accord with later theoretical discussion, in an attempt to relate divergent thinking abilities to a practical criterion of children's productive and imaginative thinking.

1.22 Construct Validity

Cronbach and Meehl (1955) observe that this term was first introduced by them and others, in the A.P.A.'s Technical Recommendations (1954), in order to help formulate the type of validity to be expected from tests that attempt to measure some attribute or quality which is not "operationally defined". The construct validity of a test is thus the extent to which the test may be said to measure some theoretical trait or construct. The construct itself however, being some postulated attribute of people, may be, in fact almost invariably is, 'bigger' than any one specific criterion. As Anastasi (1968) notes "any data throwing light on the nature of the trait under consideration and the conditions affecting its development and manifestations are grist for this validity mill".

Construct validity in fact requires the gradual accumulation of information from a variety of sources. The construct itself usually arises from psychological theory and a number of hypotheses that can eventually be proved or disproved as construct validation proceeds. This is almost precisely the way in which Guilford has built up his structure of intellect theory though as he notes (1967), even if construct validity can be built up for, say, his postulated divergent thinking abilities any practical label such as 'creative' ability applied to such factors will be a matter for validation in relation to definite indices of creative performance.

Whereas predictive and concurrent validity are examined in relation to a definite question regarding a particular empirical criterion, evidence for construct validity is not so clear-cut. As knowledge develops we arrive at greater understanding of the construct but as psychological theories change and develop so the construct also has to be modified or discarded. The process of construct validation according to Cronbach (1971), is the same as that by which scientific theories are developed, by observation, hypothesis and experiment. Reports of the thinking processes of the eminent as they engage in creative work (see for example in Ghiselin, 1955, and Koestler, 1964) lead one to hypothesise, as Guilford has done, that certain types of thinking associated with fluency and novelty of ideas are of relevance for creative thinking. The utilization of divergent thinking tests in an attempt to assess these abilities then carries the process of validation further, into the experimental stage. The construct validation of divergent thinking tests begins therefore with the question of how we can describe the thinking involved in scientific, artistic, or any other act of creation, and proceeds through interpretations of test results in relation to the construct which the tests are designed to assess.

The construct, it must be emphasised, is a theoretical one. Although there are some objectors to this non-operational aspect of validity it is a view which, Cronbach (1971) suggests, is now generally accepted. Brodbeck (1963) however argues that a construct only has scientific status when it is equated with one particular measuring operation.

This behavioural interpretation may be useful in suggesting a practical indicator of some construct but it denies the existence and usefulness of psychological concepts, such as intelligence, anxiety, or self-confidence, which arise from a whole set of theoretical and operational associations. To divorce practical measurement from psychological theory would, Cattell (1964) suggests, leave us "with a purely statistical psychometry too drily pointless to grip the attention of a psychologist with broad conceptual interests". On the other hand he also acknowledges the danger of "spurious theory, too vague and overelaborated to be useful as scientific theory".

Cronbach (1971), arguing for the theoretical nature of construct validity, notes that behaviour should be taken as an indication not a definition of the construct being investigated, and observes that no list of specific responses-to-situations, however lengthy, can define the construct, since the construct is intended to apply to situations that will arise in the future and cannot be specified now.

Cattell (1964) maintains that the term 'construct validity' may have too many empirical associations for its theoretical basis to be appreciated, and argues that 'concept validity' would be a better term in order to maintain its theoretical parentage and openness to ideational enrichment. It might result, he suggests, in a better balance between the psychological theory and the psychometric practice. Keeping the warning in mind it is acknowledged that in subsequent discussion of construct validity, the construct should be seen as a theoretical concept which exists over and above its materialisation as a set of relationships evident in empirical observation. The theoretical background to the concept of divergent thinking will be reviewed in Chapter 3.

Empirical evidence for construct validity of a test can come from many sources, from studies of performance of different types of groups on the test, from the effects of training and teaching on test scores, and from correlations with practical criteria and other tests. In fact much of the evidence accumulated from predictive and concurrent validation is relevant to construct validity if it is interpreted not in terms of the tests practical ability to predict a certain definite type of behaviour, but in terms of the psychological construct postulated as being responsible for such outcomes.

Once again it is emphasised that the same test can have validity in some, or all, of the categories discussed, each referring to a specific way in which the test might be used.

One of the major ways in which construct validity of a test is assessed is in relation to its correlations with other tests. As Campbell (1960) points out this has two aspects. We must show that the test under discussion correlates highly with other variables with which it should theoretically correlate, and also that it does not correlate significantly with variables from which it should differ. These contrasting aspects are termed 'convergent validation' and 'discriminant validation' respectively. They are very relevant to the doubts expressed earlier from Butcher (1972) regarding the correlations within batteries of various divergent thinking measures and between divergent thinking and convergent measures such as I.Q.

A great deal of evidence and debate has focussed on these issues, especially since the claims of both Getzels and Jackson (1962) and Wallach and Kogan (1966) to have established contrasting domains of 'creativity' and 'intelligence'. Burt (1962), reviewing the evidence presented by Getzels and Jackson, acknowledges that there may be a group factor of divergent thinking type abilities, but concludes that it is unreasonable to infer the existence of a separate domain outside the conception of general intelligence. In spite of the more impressive results presented by Wallach and Kogan, using a somewhat different battery of divergent tests and alternative administrative procedures, many researchers have since supported Burt's view of divergent tests, finding a limited amount of evidence for convergent and discriminant validity but also considerable overlap with general intelligence measures (e.g. Ward, 1967; Cropley, 1968; Vernon, 1972; Hargreaves and Bolton, 1972). The debate is far from settled however, conflicting interpretations even being made from the same data (Richards, 1976; Guilford, 1976) - Richards claiming that the results in question "support the Wallach-Kogan position of a unitary and independent dimensions", and Guilford that such a conclusion "is very unjustified". This issue of convergent and divergent validity will be returned to and some relevant evidence will be presented later in this study.

In studying the construct validity of a test in terms of its correlations with other tests, the technique of factor analysis has played a key role. Essentially factor analysis is a statistical procedure for locating the common factors required to account for the correlations between a number of tests. With a small number of tests and a clear pattern of intercorrelations deductions can be made without resort to more sophisticated methods, but with, for example, 20 tests and the resulting 190 intercorrelations relationships between the tests are difficult to establish. After factor analysis the intercorrelations between the tests may be accounted for in, perhaps, only five or six factors, each factor pointing to common abilities present in certain of the tests.

The interpretation of such theoretical common abilities, however, is a psychological rather than a statistical exercise and it is in the interpretation of the factors that construct validity principles emerge. Thus for example convergent validity appears if the test or tests under consideration load on a common factor with other tests thought to measure the same construct, while discriminant validity is suggested if the test does not appear significantly on the same factor as tests measuring abilities to which it is supposedly unrelated. Although it is unlikely that any test would appear in such a 'pure' a precise manner, as there are likely to be a number of minor abilities present in all tests, this method of analysis has become of central importance in modern test construction and validation. Construct validity obtained by this means is often referred to as 'factorial validity'. (Guilford, 1973.)

This is a two-edged compliment however for while it points to the important part played by factor analysis in test development and validation, it also infers the wisdom of underlining the source of such validation. There is a good deal of arbitrary judgement involved in choosing methods of factor analysis, and the test used, and the technique is not without its critics. Factor analysis has been the main tool by which Guilford has attempted to validate his structure-of-intellect model, in which "each intellectual component or factor is a unique ability that is needed to do well in a certain class of tasks or tests" (Guilford, 1959, 1968). While Guilford however claims to have demonstrated, using factor analysis, the construct validity of most of his abilities including those of divergent thinking, a number of objections have been raised (e.g. Humphreys, 1962; Cattell, 1971). Guilford's theory and some of the objections will be looked at more closely later. In general it is worth remembering that, as Thorndike and Hagan (1969) observe "this internal or 'factorial validity' still seems to need evidence of relationship to life events outside the tests themselves if the factor is to have much substance and vitality".

Construct validity will be investigated in this study both theoretically and experimentally. The theoretical basis for the construct validity of divergent thinking tests in relation to the wider issue of creative thinking will be presented in Chapter 3, and the practical investigation will involve a study of divergent performance in relation to a range of variables including intelligence and academic performance.

There are also suggestions (Butcher, 1972; Nuttall, 1972) that divergent thinking is likely to be related to personality and motivational factors as well as to cognitive ones, and the experimental study of construct validity will endeavour to include some affective as well as cognitive variables.

1.3 RELIABILITY

Test reliability refers to the consistency with which individuals score on some test. In general terms it indicates the extent to which individual differences in test scores are attributable to some 'true' differences in whatever characteristics are being assessed by the test, and the extent to which they are attributable to chance errors. More technically, measures of test reliability allow one to estimate what proportion of the total variance of test scores is due to 'error variance', and what proportion to whatever characteristic the test is able to measure consistently.

Performance on a test can be influenced by a large number of factors giving rise to variation in test scores. Following Thorndike (1951), these are commonly grouped in terms of lasting and temporary characteristics of the individual and variations due to the administration and scoring of the tests (Brown, 1976). Stanley (1971) suggests that although no list of sources of variation can be exhaustive, a consideration of such categories can help particular investigators to decide which contributions should be thought of as true variance in the quality being measured, and which as error variance. At the same time he notes that it is possible for different investigators to view the function of a test differently so that 'true variance' for one may be 'error' for another. Some investigators may, for example, treat 'test-taking attitudes' as a source of error, whilst others may consider them as characteristics which the test is meant to assess.

Although for some researchers their interest in the intercorrelations among a battery of tests at a given time leads them to adopt a measure of the 'internal' reliability or consistency of the tests at that moment in time, it is usually the degree to which a test can assess lasting characteristics that is of interest. After all, as Cattell (1964) states plainly "psychological testing has its very *raison d'être* in predicting real-life behaviour". For the purpose of prediction or the evaluation of some period of training therefore, the more meaningful definition of reliability is, as Stanley (1971) observes, "that of consistency over a comparable period of time".

No psychological test is perfectly reliable, but one must expect a high level of consistency from one performance to another or when assessed by different markers, or any subsequent claim to validity could only be of a transient and coincidental nature. Reliability in these cases is thus a prerequisite for validity, and it is frequently stated that no test can be valid unless it is reliable (e.g. Wesman, 1952, 1976). Unless the term 'reliability' has been clearly defined however this bare statement can be misleading. Wesman points out that "there is no such thing as *the* reliability coefficient for a test", and that it is relative to the situation and purpose for which it was calculated. He nevertheless includes measures of internal consistency or 'test homogeneity' within his generic use of the term reliability, though an excessively high level of homogenity can reduce the chances of the test being a good predictor of external criteria which frequently include a number of abilities. Cattell (1964) points out that "we pay a heavy price, in terms of testing inefficiency, for mistaking homogeneity for reliability. Confining the term 'reliability' to the agreement of scores 'across occasions' he notes that a high reliability is desirable, but that "an optimum rather than a high homogeneity is commonly desired". As in the case of his suggestions for a different nomenclature for the various types of validity, Cattell's proposals for a change in the titles of coefficients referring to test consistency have not been widely adopted, though his warning of the relative nature of different aspects of what is commonly termed reliability is a salutory start to this discussion of the concept. In particular it underlies the limitations in the common practice of reporting the reliability of divergent thinking tests in terms of their internal consistency rather than their consistency over a period of time. This issue and further discussion of the reliability of divergent tests will be returned to later, after an outline of the more common methods of assessing various aspects of reliability.

In the general sense in which it is used, reliability can be looked at in various ways, by re-examining with a similar test, by repeating the same test on different occasions, by examination of how individuals have performed on different sections of the test, and by the consistency with which different scorers would allocate the same mark.

1.31 Test-retest Reliability

Performance on a test can be influenced by day to day fluctuations in a person's attitude, mood and state of health. These can arise from an infinite number of changes in his personal circumstances, from having an early night to having a row at home. The degree to which test performance is affected by such random changes in the conditions of the subjects over some period of time can be assessed by comparing their performances on the same test on two different occasions. The coefficient of correlation between the scores obtained by the same individuals on the two occasions gives the 'test-retest' reliability coefficient of the test. The error variance from this method includes variance due to temporal changes and also from random changes that might occur in the testing environment, for example from changes in the weather, background noise or classroom arrangements. It also includes a source of error due to memory or practice effect introduced by repeating the test.

Examiners can reduce some of the sources of error variance by adhering to definite standards of organisation and administration that eliminate the more obvious possible sources of error such as seating arrangements, equipment, instructions, time limits and general rapport with the class. With standardised tests such conditions are clearly stated from the provision of spare pencils to standard sets of instructions. With divergent thinking tests however this is far from the case and numerous methods of administration are employed. Tests are given on a group basis with time limits for each section (e.g. Torrance, 1974), with an overall time limit for all sections (Vernon, 1971), with a mixture of timed and non-timed sections (Getzels and Jackson, 1962), and with no time limit (Cropley and Maslany, 1969).

Wallach and Kogan (1966) recommend their method of individual administration in a game-like atmosphere with no time limits.

It is noticeable that most test reliabilities arising from these situations have been internal consistency coefficients rather than test-retest coefficients. This point will be returned to later.

1.32 Alternate-form Reliability

One method of avoiding the problems of memory, and to some extent that of practice, which may affect re-test performance is to use an alternate or parallel form of the test, but maintain a time interval. This introduces other sources of variation from the use of different items in the test, but this may in fact be an advantage as measurement of error from item selection is also an important aspect in test construction. The error variance in this case is therefore due to personal and environmental fluctuations over time, and varying responses to different items in the test (internal consistency).

If an alternate form were administered without delay following the original test it would neglect to assess error due to a change in personal circumstances over an interval and be largely a measure of 'internal consistency'. Any immediate retesting however may also introduce its own errors of boredom and fatigue.

1.33 Internal Consistency

One advantage of an internal consistency measurement of reliability is that it can be assessed from a single administration of the test. This is not only convenient but also valuable when test-retest methods are inappropriate. For certain types of test the experience of taking the test once would distort the performance on a second occasion, and an internal measure of the tests consistency on a single occasion can provide valuable information about the test. It should not be taken as a substitute for reliability over time however. It assesses a different aspect of the tests consistency namely the consistency or homogeneity of individuals' performances over the different subsets or individual items of the test. It thus confines itself to internal errors arising from differing responses to test items.

In the 'split-half' method two scores are obtained for each individual, by dividing up the test into two comparable halves, consisting, for example, of the odd and even items respectively. These pairs of scores are then correlated and an estimate made of the correlation which would have been expected from the full-length test. It is known as the coefficient of split-half reliability. To take an extreme case however, if the test items were originally selected in pairs to cover some aspect of the test then an odd-even split would in effect be selecting 'matched items' of the test and would give a spuriously high estimate of the test's internal consistency over all the items.

In all cases in which the items are likely in some way to be dependent on each other so that, for instance, answering one question helps one to answer the next, internal consistency methods are not appropriate. As Cronbach (1970) notes "Non-independence gives spuriously high coefficients". This would also be true in the case of a test in which a number of questions remained unanswered, (e.g.

in a speeded test). In this case candidates would perform equally well, (i.e. score 0), on a number of items and if these were systematically, or randomly, allocated to the test 'halves' they would inflate the correlation between the halves.

A method of assessing internal consistency developed by Kuder and Richardson (1937) although still inappropriate for non-independent items or speeded tests, is preferable in many cases to the split-half method. It removes the arbitrary nature of the split-half tests where different coefficients could be obtained for different choices of halves. It is based on the performance of candidates on each item in the test in relation to all the other items, and in effect gives a reliability coefficient which is the mean of all possible split-half coefficients resulting from different splittings of a test (Cronbach, 1970).

Both methods are widely used in assessing the reliability of objective tests where a very large number of items is a common characteristic, and where problems of sampling and coverage of content are of great importance.

Their use has been extended to divergent thinking tests however and here their application is somewhat artificial. Internal consistency is nevertheless the commonest form of reliability coefficient quoted and it is almost invariably high. Wallach and Kogan (1966), for example, using the split-half method, and Cropley and Maslany (1969) using the Kuder Richardson formula 20, both reported a large number of coefficients in excess of 0.85.

Alternate-form reliabilities with time intervals however are often unsatisfactorily low. Wodtke (1963), for example, investigating the test-retest reliabilities of both verbal and non-verbal tests with alternate forms from Torrance's Creativity battery obtained reliabilities for the sub-test scores ranging from 0.03 to 0.62 for the non-verbal tests and from 0.21 to 0.74 for the verbal tests. Totalling the scores from each sub-test to give overall scores for fluency, etc., the reliabilities were slightly increased, and ranged from 0.23 to 0.63 for the non-verbal battery and from 0.35 to 0.79 for the verbal battery. Whilst acknowledging that Torrance has reported higher reliabilities with older subjects, Wodtke suggests that his results, obtained from over 100 children in both grade 4 and grade 5, do not justify the use of the tests with elementary school children, other than for research purposes.

The obvious variation in magnitude of reliabilities obtained from internal consistency and test-retest methods underlines the necessity to establish the source of any figures quoted, and to view reliability, not as a single entity but in terms of its different applications. As Stanley (1971) emphasises "There is no single, universal, and absolute reliability coefficient for a test".

1.34 Discussion

Obviously reliability coefficients from different methods are not expected to be the same as they are a measure of different sources of error. They indicate the amount of true variance that can be ascribed to the test in the circumstances being investigated. As with validity therefore, reliability has to be closely related to the purpose for which the test is being used.

The relationship between reliability and test variance can be expressed in general terms as follows. The total test variance $(\sigma_T)^2$ can be divided up into two components, true variance $(\sigma_t)^2$ and error variance $(\sigma_e)^2$, so that

$$\sigma_T^2 = \sigma_t^2 + \sigma_e^2.$$

The reliability coefficient (r_{tt}) is the proportion of total variance which is true variance, so that

$$r_{tt} = \frac{\sigma_t^2}{\sigma_T^2}.$$

Alternatively using the equation above

$$r_{tt} = \frac{\sigma_T^2 - \sigma_e^2}{\sigma_T^2}$$

i.e. it is the proportion of the total variance which remains after the proportion of error variance has been subtracted.

In modern test theory these principles are developed into general theories regarding error variance and the amount of true variance which can be attributed to variance specific to the test itself or to variance in common with other tests (common factor variance). (Harman, 1967; Guilford, 1973.) These principles will be returned to when the factor analysis results are interpreted.

To summarize the relationship between sources of error variance and the corresponding reliability coefficients, it is useful to refer to the results of some hypothetical reliability investigation with a certain (c.f. Anastasi, 1968). Suppose that alternate-form reliability after a two-month interval with 100 children gives a reliability coefficient of 0.70, split-half reliability is 0.80, and inter-scorer reliability, to be discussed later, calculated for a random sample of the scripts is 0.92. As any reliability coefficient can be interpreted directly in terms of the percentage of score variance attributable to different sources, the alternate-form reliability coefficient of 0.70 shows that 70% of the variance in test scores depends on true variance and 30% depends on error variance from sources dependent on fluctuations over time and content sampling. The split-half coefficient of 0.80 estimates error from content sampling as 20%. The difference between these estimates of error, 30% - 20% = 10%, therefore represents the error due to time fluctuations alone. The scoring error is 8%. The picture of the test as a whole can therefore be represented as follows:

True Variance	Error Variance		
62%	20%	10%	8%
	Fluctuations over time	Content sampling	Scoring variation

<u>Fig.1.</u> <u>Percentage Distribution of Variance in a Hypothetical Test</u>

The true variance taking in account all three sources of error is therefore 62%, and this represents the variance which should remain stable over changes in time, content sampling and scorer differences. All three sources of error will be considered in the present study.

The writer however doubts the appropriateness of internal consistency methods alone when applied to divergent thinking tests and the alternate-form method will be preferred to cover fluctuations over time and test content. Guilford

(1973) points out that reliability based on internal consistency "tells us nothing about function stability of persons or of tests" though he also notes that "it does have meaning in connection with factorial descriptions of tests". High internal consistency tells us that whatever a test is measuring at a particular time it is doing so consistently throughout the test and there is little error variance due to the choice of items. If as in factor analysis our interest is confined to the relationships between results of particular tests at some particular time, this type of internal consistency is appropriate. It is rather like Hudson's comment in justifying the use of open-ended tests "not in terms of their test-retest reliability", but in terms of their power to differentiate between other variables. As noted earlier whether these differences are pervasive is a different question needing concurrent and predictive validation. In such cases test-retest or alternate-form reliability is needed to examine the tests reliability over time, for a test which has low reliability due to variable errors over some interval must also have low validity. It cannot correlate, except by change, more highly with a criterion than with itself.

It is suggested that as our use of divergent thinking tests extends into predictions and into use as criteria for evaluating teaching methods and curricula content, more attention should be given to reliability over time. Although interest in divergent thinking tests has moved ahead into questions of what can be inferred from the results about individuals' future performances, reliability has too often remained at the stage of internal consistency.

The goals of validity and reliability are often difficult to reconcile, for as Guilford (1973) observes, "when we seek to make a single test both highly reliable (internally) and also highly valid, we are often working at cross purposes...... In aiming for one goal, we may defeat efforts towards the other". If a test is highly homogeneous it would have a high internal reliability but it might correlate with little else but itself or very similar tests. Its validity might be defended in terms of construct validity of the discriminant or convergent type, but it would need a very closely related criterion to show appreciable correlation in other forms of validity.

So far a large proportion of the claims made for the validity of divergent thinking tests has been made on this basis. Wallach and Kogan (1966) for example emphasising the internal consistency of their tests and their lack of correlation with intelligence measures. While, as noted earlier, others have been less successful than Wallach and Kogan, a lack of support for a domain of divergent thinking or 'creativity' quite distinct from intelligence and other measures would not negate the value of such tests, provided they have some element of validity in recognising thinking abilities that are not commonly measured by more conventional tests. As Vernon (1967) suggests, they may, in pointing to a new area of abilities, help "overcome the deadening influence of assessing the growing generation for conformity".

Partly, the writer suggests, in order to accommodate internal consistency methods however some divergent thinking tests have been composed of a number of items (e.g. Wallach and Kogan, 1966; Ogilvie, 1974), when only one or two items would have sufficed. It could be argued, in fact, that the latter would be more likely to yield a meaningful measure of divergent thinking. Wallach and Kogan themselves argue that children should be given plenty of time to come up with ideas, for, "if responses of greater stereotypy are likely to come early in a sequence even in the case of creative persons, then it will not be possible to detect these persons if insufficient time is permitted". They also argue for a form of administration that is "relatively free from the stress of knowing that one's behaviour is under close evaluation". They consequently administer the divergent thinking exercises in a game-like atmosphere on an individual basis with no time limit. However having conducted the "Uses for objects" test by asking for "all the different ways you could use a newspaper" they repeat the exercise seven more times with the stimulus object being a knife, automobile tyre, cork, shoe, button, key and chair. In Wallach and Kogan's experiment the atmosphere might have been sufficient to maintain children's motivation to really think hard throughout all the items, but it is suggested that particularly in a written group presentation (e.g. Ogilvie, 1974), the number of items is likely to result in pupils doing less than justice to their divergent thinking ability. Instead they are likely to spread their efforts evenly over the items available without entering deeply into the type of thinking in which, as Guilford (1959, 1973) describes divergent thinking, "considerable searching about is done". The result will obviously be high internal consistency though, it is suggested, at the expense of validity.

Objections to the interpretation of the results of divergent thinking tests as being independent of general intelligence can be made on the grounds that such independence can often be explained by their low reliability. To increase reliability by increasing the number of items, however, is likely to be unprofitable in the long run, when such tests are subjected to criterion-related validity studies. It is also questionable whether the items are of sufficient independence to allow the application of statistical measures of internal consistency.

It is noticeable that Torrance (1974) in the Manual for his divergent thinking tests does not refer to internal consistency in commenting on their reliability, nor does he use multiple items in his individual tests. He claims instead some test-retest reliability for his tests, over various periods of time, the longest being after a period of almost three years. The evidence is varied, though the majority of the retest reliabilities quoted were higher than those given by Wodtke (see earlier page 29) and lie in the reasonably satisfactory region of 0.60 to 0.70. Most of these, however, arose from cumulative scores rather than from individual tests and the reliabilities would consequently be inflated. Torrance's disappointment at some of the results is nevertheless apparent in his remark that the motivational aspects of the testing situations may not have been maintained and since the latter is indisputably important in creative thinking reliability would be affected.

Apart from differences in administrative conditions due to chance factors there are also wide differences in the actual procedures recommended by different researchers. Since Wallach and Kogan's emphasis on individual and informal testing procedures a number of researchers have investigated the effects of different methods of administration. Results have been conflicting, relaxed conditions, for example, appearing to be more favourable in some experiments (e.g. Vernon, 1971) and 'mild stress' (Channon, 1974) in others. Time allowances have been seen as exerting less influence than other characteristics of the test situation (Hargreaves, 1974) and also as the most potent factor (Van Mondfrans *et al*, 1971). Direct comparison of the various studies is difficult however as superficially similar testing batteries and types of administration are often composed very differently, informal procedures, for example, varying from allowing the tests to be taken home to answer at leisure to attempts at establishing a game-like atmosphere within the classroom.

What does appear from most of the investigations however is confirmation of the presence of different motivational effects. This underlines the importance of arriving at some set of optimal conditions that will allow the pupils to do justice to their divergent thinking abilities and yet be sufficiently standardised to be replicable by other investigators without undue variation. No direct investigation is being made into differing procedures in this study, though some relevant observations will be made when appropriate. In general the tests will be administered under relaxed but timed conditions after the experimenter has become well acquainted with the children concerned. Details will be given later.

On the basis of this discussion the tests in the study will be based, in a similar manner to Torrance, on single items only; and to cover fluctuations over time and test content reliability will be assessed by the alternate-form method after an interval of one week. Details of the tests and the form of administration will be given later.

1.35 Long-term Stability

Reliability coefficients obtained by a retest or alternate-form method after an interval of time are sometimes referred to as coefficients of 'stability', as opposed to the coefficients of 'consistency' arrived at by internal measures. The term is perhaps better reserved however for long term studies of the constancy or stability of some ability, over a considerable period of time rather than for shorter periods when it is the reliability of the test rather than the trait that is then being investigated. While studies of retest reliability are generally restricted to short-range random changes that occur within a few weeks of the initial testing, questions of stability can extend for some years after the initial testing.

Cronbach (1970) quotes several retest correlations for the Stanford-Binet scale. Over about a week the 'retest correlation' for 200 seven-year-old children was 0.91, over 4 years it was 0.74 and over 11 years, 0.68. Without entering into discussion of the possible self-fulfilling nature of the children's education between 7 and 18 a high proportion of the test variance at age 7 can still be measured at age 18, or looked at from the opposite point of view a considerable proportion is not stable over time.

If, however, education places some value on fostering divergent thinking in the belief that it will be relevant to future work, it is important to have some information about the stability of the trait. Haddon and Lytton (1971), for example, in one of the very few longitudinal studies of divergent abilities, suggest that promotion of divergent thinking abilities in the primary school still has a significant effect on performance in divergent thinking tests four years later. Whether children's divergent thinking abilities increase with age or whether education has a conforming influence resulting in diminishing levels of divergent ability is also a matter of long-term study.

Some evidence on these issues of long-term stability and the development of divergent thinking abilities will be presented in this study. This will be produced by following up after an interval of five years, a large proportion of the 265, 11 year-old children involved in an earlier study of the writer (Richards, 1970). The results should have considerable implications for the use of divergent thinking tests as predictive measures.

1.36 Scorer Reliability

Psychological tests designed for group use are usually provided with highly standardised procedures for scoring and the scorer has only to follow carefully the instructions given. In such tests there is consequently very little error variance attributable to scorer differences. In scoring divergent thinking tests however a good deal is left to the judgement of the scorer and considerable variations could occur. There are no 'right' answers to divergent tests, testees are encouraged to be unusual and inventive and scoring procedures must attempt to credit such responses without departing entirely from judging the appropriateness of the anwers.

In these circumstances it would not be surprising if two independent scorers arrived at somewhat different results. Consequently, as Anastasi (1968) points out, with this type of test "there is as much need for a measure of scorer reliability as there is for the more usual reliability coefficients", and "it is imperative to ascertain scorer reliability for all such instruments". In spite of this warning however the caution that one would expect from users of such tests has not been very apparent. Even details of the scoring procedures are rarely given and measures of inter-scorer reliability are rarer still.

Wallach and Kogan (1966) commenting adversely on aspects of the Getzels and Jackson study (1962), note that "we might note such issues as the problem of scoring reliability" and observe that for certain measures "no descriptions of scoring methods nor information as to reliability of scoring were provided". Such criticisms could be levelled at a large number of studies involving divergent thinking, yet the question of scorer reliability is of crucial importance for tests where no completely objective method of scoring is possible.

Scorer reliability is found by having a sample of scripts independently scored by two markers. The two scores thus obtained for each script are then correlated and the resulting correlation coefficient is a measure of inter-scorer reliability. The amount of error from this source is distinct from that arrived at by retest and internal consistency measures and can be represented as shown earlier

(Fig.1, page 32). The chief reason why inter-scorer reliability is seldom reported is no doubt due to the time and effort needed to score divergent thinking tests. For scoring anything more than fluency the procedures are quite lengthy and, unlike scoring standardised tests, potential scorers also need to be aware of the principles underlying the tests.

Vernon (1971) observes that "a preliminary difficulty to be overcome in using divergent thinking tests is the complexity and tediousness of scoring, though this is seldom referred to in the literature", and Cronbach (1970) points out simply that "scoring tests of this sort is difficult". Questioning the relationship some researchers have found between age and divergent production, Cronbach also suggests that with a subjective element in the scoring procedures, scorers may inadvertently be more severe in scoring work of older persons. The writer suggests that subjective impressions can also seriously affect one's assessments of children within groups.

Some time will be given in this study to investigate this possibility. Details of scoring procedures from other studies will be reviewed, and every effort made to establish reliable scoring procedures for present use, and possibly for further investigations by the writer or others. Without such procedures any subsequent results regarding the reliability and validity of the tests could only be of a personal and fortuitous nature.

1.4 SUMMARY

It has been suggested that many doubts exist about the reliability and validity of divergent thinking tests, not the least in relation to their label as 'creativity' tests. Neither reliability nor validity is an absolute concept however and the previous discussion has attempted to clarify some of their different aspects, and their application to tests of divergent thinking.

The action to be taken in the present study to investigate the issues raised involves investigation of the following:-

(a) Scoring procedures and inter-scorer reliability.

(b) Alternate-form reliability over an interval of one week.

(c) Long-term stability over an interval of five years.

(d) Concurrent validity, including ratings of children's creative interests and behaviour, and a practical exercise designed to entail divergent thinking.

(e) Construct validity in relation to a range of variables both cognitive and affective.

The theoretical basis for construct validity will also be discussed (Chapter 3) and practical investigations of reliability and validity will be reviewed (Chapter 4).

The next chapter will look more closely at two issues remaining outstanding since attention was drawn to them in the Introduction. Firstly, the educational relevance of a study of creative abilities and what has been disparagingly referred to as the "cult of creativity" (White, 1968), and secondly, the relationship between divergent thinking tests and so-called tests of creativity. In order to clarify the discussion some time will also be spent on the problem of defining creativity.

CHAPTER TWO

CREATIVITY, DIVERGENT THINKING AND CLASSROOM LEARNING

2.1 CREATIVITY AND THE DEVELOPMENT OF CREATIVE THINKING ABILITIES IN THE CLASSROOM

A large proportion of the creativity research that has been related to children's thinking and school learning acknowledges the early work of Torrance (1959, 1962, 1963) in identifying and developing creative abilities in the classroom. Much of this research involves Torrance's divergent thinking tests, (labelled the Torrance Tests of Creative Thinking, latest version: Torrance, 1974) which he has adapted from the Guilford material, for use in schools from "kindergarten thru graduate school".

In addition to his influence in designing divergent tests to assess children's creative thinking abilities, Torrance has also played a significant part, in books such as Guiding Creative Talent (1962) and Rewarding Creative Behaviour (1965), in influencing approaches to the school curriculum to develop such abilities, and he emphasises the part education should play in developing these "creative characteristics", both for the individual's personal fulfilment and to satisfy the needs of society for creative initiative. He considers a formal, passive, authoritarian approach to learning to be unlikely to stimulate a child's capacity to think creatively. Instead he suggests that "a child learns creatively by questioning, inquiring, searching, manipulating, experimenting, even by aimless play; in short, by always trying to get at the truth" (Torrance, 1963).

Crutchfield (1964) also emphasises the freedom needed for children to develop their creative talents, and suggests that pressures on a child to conform inhibits his creativity and quells his motivation so that he loses his "trust in the essential validity of his own processes of thought and imagination". Mooney (1967) commenting on 'Creation in the Classroom setting', underlines the importance of communication between teacher and pupil in a spirit of mutual enquiry, and Wallach and Kogan (1966) make a similar plea for classroom conditions conducive to fostering creativity, and advocate a stress-free atmosphere in which the creative child "can blossom forth cognitively".

Though such assumptions are likely to be tempered somewhat under the "astringent intellectual scrutiny" which the Plowden Report (1967) recommends that teachers should bring to bear on their day to day work, and by warnings of indiscriminate reaction against values of the traditional school curriculum (Dearden, 1968; Peters, 1969), they express a philosophy which underlies much of the curriculum development that has taken place both in this country and in America in the last twenty years. In particular the word 'creativity' has become closely linked to teaching methods that emphasise the role of experience and personal discovery in children's learning.

Critics of this link however have suggested that its dependance on a personal 'everyday' concept of creativity is becoming confusing, meaningless and even anti-educational (White, 1968), and that mere 'openness' may in fact lead to the child's level of adjustment being less than adequate to meet the demands of the adult world (Shouksmith, 1970). If the word 'creative' is used to describe any activity instigated by the child himself, criticisms are understandable, for it might well appear that at classroom level "one only need speak to be creative" (Dearden, 1968) or that creative answers appear to be equated with the ability to write (Hudson, 1966).

To be able to think of new ideas and to act for oneself however are abilities to be encouraged in children, and methods of fostering them in the classroom deserve attention. The real danger is that means may be mistaken for ends especially if they are given the title of 'creative' activities. Partly because of the conceptual confusion surrounding such a term, Ausubel (1968) suggests that "many otherwise hardheaded educators have adopted highly unrealistic educational objectives regarding the nurturance of creativity, and many otherwise well-trained educational and school psychologists have deluded themselves into believing that they are able to identify pupils with unusual potentialities for creativity".

Accepting the desirability of 'teaching for creativity' is to adopt an attractive principle, but it is a long way from knowing how to put it into practice. The questions of how to recognise, and how best to foster creative abilities, are still far from having any definite answers. One step in the right direction however is to attempt to clarify what is meant by creativity.

2.11 Defining Creativity

In studies of creativity a rough distinction can be made between those orientated to products, persons or processes (Brogen and Sprecher, 1964; Wallach, 1970). The first two of these perspectives normally focus on adult populations, and an extensive literature (e.g. Roe, 1953; Cattell and Drevdahl, 1955; MacKinnon, 1962; Taylor C.W. *et al*, 1963; Taylor D.W., 1963; Cross *et al*, 1967) has accumulated regarding the occupational achievement and personalities of individuals in both the arts and sciences.

Product-centred criteria however have been extremely varied in nature, including indices such as salary, number of publications or patents, membership of professional bodies and ratings by peers or panels of judges, and it is not surprising that Golann (1963) finds that they turn out to be relatively uncorrelated among themselves. The creative person too is revealed as a very complicated person, MacKinnon (1962) for example finding from his study of creative architects that "the successful and effective architect must, with the skill of a juggler, combine, reconcile, and exercise the diverse skills of businessman, lawyer, artist, engineer, and advertising man, as well as those of author and journalist, psychiatrist, educator, and psychologist"!

It is on the third aspect, that of processes that research involving divergent thinking tests has concentrated in an effort to recognise and foster psychological processes that may be more generally related to creative thinking abilities than an emphasis on products or persons.

Taylor (1964) is critical of a process definition of creativity preferring to rely on evaluation of the product. He suggests however that no single definition of creativity will suit all workers in the field, and he is no doubt right in suggesting that creativity is likely to be a complex rather than a unitary concept. While some researchers would consequently adopt different labels for various abilities thought likely to be of value in creative production, divergent thinking being one of them, Taylor recommends the adoption of different definitions of creativity! He encourages a blanket use of the word to enable researchers to move ahead in their work either by choosing "tentatively an existing definition of creativity" or by developing "a definition of their own". This 'inspirational' use of the term no doubt helped encourage the early enthusiasm for creativity research in such gatherings as Taylor's 'University of Utah Research Conferences on the Identification of Creative Scientific Talent' which began in 1956, though subsequent evaluation of the researches is made almost unmanageable by the diversity of definitions. Some of the confusion is removed if the particular context of the definition is made clear, but even this is not always done and readers of the research literature consequently tend to interpret the terminology in their own way, not necessarily in the way the investigators intended.

The definitions of creativity favoured by the organisers of the Utah conferences were directed towards the evaluation of the degree of creativeness of a product or a performance as distinct, Taylor (1964) notes, from estimates of potential such as test scores. Taylor recommends two definitions, that of Ghiselin (1957), "that the measure of a creative product be the extent to which it restructures our universe of understanding", and that of Lacklen R.L., who reported at the Second Utah Conference (Taylor, 1958) that in scientific work at the Space Agency creativity was defined by "the extent of the area of science that the contribution underlies".

Taylor suggests that despite certain differences in emphasis the two definitions were considered to be quite compatible, and showed the same emphasis on the creative product rather than creative process. Whilst recognising that indicators of creative potential can be assessed from both products and processes he suggests that criteria for the evaluation of degree of creativity are much better sought in the product rather than in assessment of the process.

With populations of research scientists or in evaluating the contributions of exceptional individuals Taylor's considerations are no doubt important and a good deal of creativity literature is concerned with establishing criteria for judging the creative products of talented individuals. In any absolute sense however, what is creative has to be debated in terms of the values, judgements and stage of development of a particular culture (Dentler and Mackler, 1962), and with varying aesthetic and professional points of view it is not surprising that the problem of evaluating creativity in terms of products has not reached any consensus. Golann (1963) claims that criteria that involve evaluating the product have been found to be so inconsistent that they are virtually useless, and Shapiro (1968) observes that one of the disheartening conclusions emerging from approximately fifteen years of intensive research is that little progress has been made on achieving acceptable criteria of creativity.

Welsh (1973) is similarly dismayed at the lack of success in arriving at an acceptable definition. Despite the increasing attention being paid to creativity he suggests that "it does not seem possible to offer a simple, substantive definition of creativity that would win consensus. This is borne out by most reviewers of the field. Taylor I.A. (1959) analyses over one hundred different definitions, Dudek (1974) somewhat belatedly notes that "there are some fifty definitions of creativity", and Dacey and Madeus (1969) more realistically point out that the number of definitions of creativity are "only slightly fewer than the numerous works on the subject".

Any attempt to establish a consensus definition of creativity is therefore unlikely to succeed, and it underlines the deceptive use of the word in relation to divergent thinking test alone. At the same time in order to move nearer towards an understanding of creativity in its totality we must, as Welsh (1973) suggests, first understand some of the bits and pieces of human behaviour related to the subject "regardless of the fact that the observed behaviour may seem at this time to be far removed from the supernatural acts of a creative genius".

Different perspectives within the broad study of creativity are each likely to be important contributions to this understanding, provided investigators refrain from assuming that the concept of creativity can be conceptualised fully within any particular perspective. It is suggested however that investogations of creativity as a process may be more profitable than other apparatus, not only in identifying and developing those abilities that are considered likely to enhance one's potential for creative production, but also in developing practical criteria related to such a process. Rogers (1954, 1970) defines creativity as a process resulting in "emergence in action of a novel relational product, growing out of the uniqueness of the individual on the one hand, and the materials, events, people or circumstances of his life on the other". This allows him to relate the concept of creativity to a personal level as Bruner (1960, 1961) does for 'discovery' and Guilford (1950) for 'creative abilities'. Bruner (1960) in the context of discussing school learning, makes it clear that the concept of discovery should not be restricted to the act of finding out something that before was unknown to mankind, but rather should relate to the individual, and to all forms of obtaining knowledge for oneself by the use of one's mind.

Guilford (1950) discussing creative abilities maintains that the general psychological conviction is that "all individuals possess to some degree all abilities, except for the occurrence of pathologies". Whatever, therefore, the components of creativity might be this principle of 'continuity of abilities' maintains that they are present to some extent in all of us. In this way, a study of 'creativity' in terms of a process rather than a product can provide a basis for investigating and developing creative abilities in all individuals not just in those who have distinguished themselves in some field. In a similar way, Rogers (1954, 1970) sees the creative process as being present, to some degree, in most activity so that "the action of the child inventing a new game with his playmates, Einstein formulating a theory of relativity, ... all of these are, in terms of our definition creative".

In order to study creative behaviour, Torrance (1974) is also convinced that the definition of creativity should be based on that of a process:-

> "a process of becoming sensitive to problems, deficiencies, gaps in knowledge, missing elements, disharmonies, and so on; identifying the difficulty; searching for solutions, making guesses, or formulating hypotheses about the deficiencies; testing and retesting these hypotheses and possibly modifying and retesting them; and finally communicating the results." Torrance (1974).

Torrance sees this definition as describing a natural human process that it is in harmony with historic usage and equally applicable in scientific, artistic, literary, dramatic, and interpersonal creativity. It is in fact such a comprehensive definition that it reminds one of Wechsler's (1958, 1974) definition of intelligence as "the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment", and his comments that "it is global because it characterises the individual's behaviour as a whole; it is an aggregate because it is composed of elements or abilities which, though not entirely independent, are qualitatively differentiable". Torrance's definition, on which he has based his whole work on creativity, is seen by Ausubel (1963) as defining "a generalised constellation of intellectual abilities, personality variables, and problem-solving traits", and while he does not deny the existence of such a general group of creative abilities he objects to Torrance's labelling of the abilities as if they were a highly particularised and substantive capacity.

This objection however is levelled less at Torrance's definition than at his operational interpretation of the definition in terms of his creativity test. The limitations of the latter will be discussed in the next section. For the present it is suggested that Torrance's global definition of creativity has much to commend it. Guilford (1950) appeals to a similar general concept of creativity in terms of "those patterns of traits that are characteristic of creative persons", and like Ausubel he sees these as including a broad category of abilities, applitudes, interests and temperamental qualities.

It is important to realise that process definitions of creativity are based on a person's 'potential' for creative performance in real life not on the actual quality of his output. Torrance (1974) points out that a high level of the abilities in his process definition do not guarantee that the possessor will behave in a creative manner, rather, they increase his chance of behaving creatively.

Guilford similarly emphasises that creative abilities determine whether the individual has the "power to exhibit creative behaviour to a noteworthy degree". Whether or not the gifted individual actually produces results of a creative nature depends on other factors including motivation and opportunity.

Drevdahl (1956) in an attempt to define the creative process also finds himself formulating a generalised definition of creativity as:-

> "the capacity of persons to produce compositions, products, or ideas of any sort which are essentially new or novel, and previously unknown to the producer. It can be imaginative activity, or thought synthesis, where the product is not a mere summation. It may involve the forming of new patterns and combinations of information derived from past experience, and the transplanting of old relationships to new correlates. It must be purposeful or goal directed, not mere idle

> fantasy - although it need not have immediate practical application or be a perfect and complete product. It may take the form of an artistic, literary or scientific production or may be of a procedural or methodological nature."

While this type of generalised definition based on creativity as a process appears to be one way of incorporating all the different aspects of creative ability, its very generality means that in practice it verges on the meaningless. In principle however it reminds one that creativity is not a specific term and that investigations of children's behaviour within the overall field should be related not to different, more specific, definitions of 'creativity' but to other more specific and more easily defined variables.

One way of narrowing the field has been to study the types of thinking that appear to be involved in the creative process. It should be emphasised however that to label any one particular type of thinking as creative is likely to result, as in attempts to categorise the 'creative product', in a multitude of different thinking abilities claiming to justify the title. As Vernon (1973) points out, "to the psychologist ......, creative thinking is merely one of the many kinds of thinking which range from autistic fantasy and dreaming to logical reasoning". "Indeed," as he adds, "to some extent it seems to partake of both extremes".

Many writers however have attempted to specify criteria for creative thinking usually associated with novel or unusual combinations of ideas. Thurstone (1952) sees creative thinking in terms of what is novel for an individual and argues that it makes no difference how society regards the idea. Stein (1967) on the other hand claims that thinking can only be called creative if it results in a product which is not simply novel to the individual but obeys certain external criteria which are relative to the society of which the individual is part. This, however, is to return to the difficulty of defining the creative product, and it is suggested that specific definitions of creative thinking are likely to be similarly constricting and in the long run confusing. Instead, in parallel to the definition of creativity as a general process, it is suggested that creative thinking should be seen as a multifaceted activity. It would then be more appropriate to talk in terms of creative thinking abilities and to designate any particular ability in some specific way. In general therefore creative thinking will be concerned with *a process resulting in some new thought, insight or material product; the difference between what is new to an individual and what is judged as valuable and original by Society at large being only a matter of degree.* In this way thinking abilities that have variously been labelled "productive" (Wertheuner, 1961), "adventurous" (Bartlett, 1958), "lateral" (De Bono, 1967) or "divergent" (Guilford, 1956) can be recognised as being relevant to creative thinking in general.

2.12 Educational Implications

With creativity defined in terms of a process involving a variety of creative thinking abilities, it is possible to consider means whereby education might foster such abilities, and it enables teaching to be more directly related to developing children's creative thinking not simply as a vague ideal but with particular thinking skills in mind.

Torrance (1972) maintains that improvements in creative thinking skills are most predictable when direct teaching is involved, and Bruner (1961) emphasises the part that "inductive teaching" can play in helping pupils to be exploratory and constructive in their thinking. Even Burt (1962), while critical of any concept of creativity confined to divergent abilities, welcomes the attention that schools might give to individuals whose special abilities are of a productive rather than a merely reproductive type.

Wallach and Kogan (1966) after emphasising the importance of divergent abilities nevertheless acknowledge that there are both divergent and convergent aspects of creative thinking and that both should be developed in the classroom. They suggest that more attention should be given to problem solving and personal exploration as each provides opportunities for fostering associative and inferential modes of thinking.

More specifically de Bono (1973) has created materials specially designed for "teaching thinking directly". In the past, he suggests, education has emphasised thinking skills concerned with the development and utilization of ideas rather than with the making of new ideas, and he emphasises the need to develop what he terms "lateral thinking" and "the more creative aspects of mind" (de Bono, 1969).

In America Wootton (1965) describing the work of the School Mathematics Study Group notes that many of the exercises were designed to be of a discovery type to extend the treatment in the text and also "promote original thinking and creativity on the part of the student", and similar beliefs in the value of exploratory material to encourage children's thinking processes are apparent in a number of British projects. The Nuffield Primary Mathematics Project aims to foster in children "a critical, logical, but also creative turn of mind" Nuffield Foundation (1964a), and the director (Matthews, 1969) notes that the central message of the project was to "let the children think". A further Nuffield Project (1964b) was designed to "encourage children to think freely and courageously about science", and Kerr (1966) suggests that the central purpose of the Nuffield science courses was "to stimulate critical and imaginative thinking".

It is noteworthy that the above suggestions for encouraging creativity do so, not in the trivial sense of any spontaneous action of the child, but in terms of those original and imaginative aspects of children's thinking that are likely to contribute to the process of creative thinking. Although the implementation of some of the suggestions might have been due to the mood of the times and the "wave of romantic enthusiasm for the spontaneous" that Butcher (1972) maintains characterises some of the interest in creativity, the aims mentioned are firmly based in the realm of children's creative thinking abilities, and show a considerable overlap with the views of Guilford (1958) relating divergent thinking skills to "creative thinking and invention".

Not all the aims are easy to translate into definite teaching objectives however, and although they clearly interpret creativity in terms of creative thinking abilities they could look more closely at the skills involved so that the teaching could be more specifically directed towards those ends.

Shouksmith (1970) supports the view that creative thinking abilities can be encouraged and developed, but also suggests that educationalists in general have taken an altogether too simple view of creativity, and that to encourage children to express themselves openly is not enough. He considers instead that children should be educated in the use of appropriate strategies to meet a wide range of problem situations. This view is in accord with Guilford's (1952) suggestions that like most behaviour creative activity probably represents to some extent many learned skills, and that appropriate teaching methods and techniques will help promote such abilities including those of divergent thinking (Guilford, 1967).

Torrance (1972) reviewing attempts at improving children's creative abilities by using specific teaching programmes reports that improvement was shown on over 90% of the criteria used. As the latter were largely divergent thinking tests the effectiveness of the teaching methods in improving other thinking abilities should be questioned, but looked at in terms of divergent abilities alone the improvements are consistent with Guilford's view that such abilities can be enhanced by appropriate teaching. Based on a similar view held by the writer, further support for this finding is reported by Franklin and Richards (1977) as the result of teaching procedures designed specifically to foster children's divergent thinking abilities. The latter avoids the issue of whether divergent thinking tests justify their common title of 'creativity tests', but this question is now considered in the next section.

2.2 DIVERGENT THINKING TESTS - NOT 'CREATIVITY' TESTS

Since Guilford's (1950) re-emphasis of the term 'creativity' many psychologists have adopted the word to describe divergent thinking tests in much the same way as educationalists have been tempted to equate 'creativity' with 'freedom of expression' in the classroom. To suggest, Brown (1976), that in order for a behaviour to be indicative of creativity, or a test classed as a measure of creativity, responses must be produced rather than just selected from among available responses; is to adopt a narrow and limited interpretation of both creativity and divergent thinking.

Even writers who accept a wide interpretation of creativity in terms of the process definition suggested in the last section argue for the retention of the word creative to label divergent thinking aspects of creative ability. Even if, as Wallach and Kogan (1966) suggest their divergent thinking measure, based on the production of fluent and novel associates related to some stimulus, can be shown to be "strikingly independent of the conventional realm of general intelligence, while at the same time being a unitary and pervasive dimension of individual differences in its own right", it seems highly undesirable for such a dimension to be labelled 'creativity'. They do not themselves deny the value of other abilities, particularly general intelligence, in achieving real-life creative performance, but their limited use of the term 'creativity' with its exclusion of abilities like intelligence may well be anti-educational if it were to be taken by teachers and others to be more than just a label for one type of psychological ability - no matter how important those particular abilities might be.

It is therefore of some concern to the writer that the tests have been widely used and described as criteria of creativity in evaluations of curricula and teaching approaches. This is particularly common in America but also not unusual in this country, Ogilvie (1974) and Starr and Nicholl (1975), for example, both accept the term creativity as describing an activity based on divergent thinking. Torrance (1972) reviewed 142 studies of programmes designed to "teach children to think creatively", 103 of which used performance on his divergent thinking battery, described as 'Tests of Creative Thinking' as criteria. The latter ranged from evaluations of children's creativity due to 'creative versus non creative school climate' (Raina, 1971), 'multimedia sensory exercises' (Abbott, 1972) and 'music improvisation' (Vaughan and Myers, 1971), to teaching approaches based on programmes of 'creative expression' (Schaefer, 1970) and 'creative problem solving' (Olkin, 1967). Notwithstanding this diversity, Torrance's tests were often the sole criterion of creativity used.

It appears that the 'bandwaggon' of creativity research remarked on by Hudson (1966) is now running markedly out of control. It is certainly running away with much of the validity that Torrance might claim for his tests if they were more precisely defined in terms of divergent thinking. While he notes that "I strongly favour and have used more 'real life' criteria" Torrance (1972), he nevertheless defends the validity of his tests as tests of creativity, in the belief that "the word describes the behaviour investigated more adequately than any other word I know".

Torrance describes the creative behaviour being investigated in terms of his process definition given in the last section (page 52) but while his definition can be justified as a global definition of creativity his claim that the behaviour is adequately assessed by his divergent thinking measures is asking a great deal more than appears realistic. He modifies his claim somewhat by mentioning that it is people's potential for creative behaviour that is assessed, but even so the abilities assessed by his tests would seem to sample only some of the creative thinking abilities that could be suggested.

Even though Guilford (1965) notes that he believes his divergent thinking category "to contain some of the most directly relevant intellectual abilities for creative thinking and creative production", it must be emphasised that he does not equate divergent thinking with creativity. His structure of intellect model, to be discussed later, contains a wide range of other intellectual abilities including those covered by the conventional type of I.Q. test, and though critical of a conception of creative talent accounted for solely in terms of high intelligence or I.Q., - a concept he suggests is largely responsible for the lack of progress in the understanding of creative people - he emphasises that creativity is likely to be a matter involving the entire personality "including intelligence".

Torrance on the other hand, whilst accepting that creative abilities and tendencies are likely to be related to a 'constellation of general abilities', sees these abilities largely in terms of divergent production. He notes that his tests involve different kinds of thinking that incorporate many features from models of the creative process but also that "An attempt is made, however, to assess the products that result from the administration of these test activities in terms of Guilford's divergent thinking factors (fluency, flexibility, originality, and elaboration)" (Torrance, 1966). It is in this step from his general definition of creativity to the specific abilities covered by his tests that Torrance's theoretical arguments become separated from his practical assessments. Notwithstanding this he labels the tests, in association with his process definition of creativity as tests of 'Creative Thinking'.

Taking Torrance's test in conjunction with the wide use of other batteries of divergent thinking tests designed or adapted for their own use by other researchers, including Getzels and Jackson (1962) and Wallach and Kogan (1966) but also labelled as 'creativity' tests, a very large proportion of 'creativity' research needs to be put in perspective as being largely an investigation of divergent thinking abilities.

As Torrance (1966) observes in commenting on the development of his tests, they probably represent the first time that a battery of such tests has ever been made available for general research use, at least with children. Their wide classroom use however is another source of concern, for teachers are likely to be less aware and less critical than researchers of their limited validity in relation to real-life creative performance or the whole spectrum of creative thinking abilities.

Getzels and Jackson (1962) justify their use of the 'creativity' label by noting that the tests require the pupils to 'diverge' in their responses so that they measure "the ability to deal inventively with the verbal or numerical symbols or with object-space relations". But as Burt (1962) points out creativity should imply "useful creative activities", and in order to deal rationally as well as inventively with material supplied he suggests general intelligence is an essential constituent. Most investigators would agree with this assessment, but the labelling of divergent activities as 'creativity' is still not always avoided.

Cropley (1967) writing on 'creativity', notes that "virtually all the research cited in this book, whether the authors reported the results as concerning divergent thinking or creativity, employed tests of this latter kind in order to identify highly creative children".

While pointint out the "doubtful validity" of this practice he nevertheless suggests that although the "concept of creativity is a difficult one to employ with precision because of its impreciseness, the term is coming to be accepted by many psychologists and educators as referring to an intellective mode characterised by thinking of the divergent kind". He therefore argues for an 'objective' use of the term creativity in a way which will allow it and the term divergent thinking "to be used almost interchangeably". He points out that, personally, he would prefer to use the notion of divergent thinking exclusively, in order to avoid any possibility of confusion between the everyday conception of what 'creativity' means and a strict psychometric meaning of the term as a shorthand for referring to the fact that certain tests (of divergent thinking) go together. He concludes, however, somewhat contradictingly, that "the term 'creativity' is a very meaningful one, and is by now so well established in the literature and so widely used that strict insistence on the exclusive use of 'divergent thinking' would involve restating the findings of almost all research in the area; and would be unnecessarily pedantic in the present context". He suggests that if the reader remembers that there is no evidence that a Michaelangelo or Einstein would have done well on these tests, it is legitimate to refer to them generically as 'creativity tests'.

Most researchers would agree with Cropley's general assessment of the 'creativity' label as being synonymous with that of 'divergent' in the context of research based on divergent thinking tests, but there are many, including the writer, who would not wish to perpetrate the problem by continuing to apply the 'creativity' description to divergent tests. Cropley's answer, to remind readers that 'creativity' is 'what divergent thinking tests measure' is almost a parallel with the traditional operational definition of intelligence and seems to be walking into the same trap. If one continues to use the label 'creativity' to describe certain tests it is irresponsible to assume that the man in the street will interpret the label in a strict psychometric sense. If psychologists and educators continue to use the label it is feasible to suppose that a large proportion of people will assume that such a trait exists. If on the other hand, being more sceptical, they were to scrutinise the label more closely, they might indeed be critical of a 'cult of creativity' and their disillusionment might understandably alienate them from results which, expressed more accurately might have valid and important implications.

Thankfully the label 'creativity' to describe divergent thinking tests is being applied less frequently, particularly in this country. Vernon (1971) for example, in an investigation of divergent thinking tests refers to them consistently as such, only mentioning the "area of divergent thinking" as that of "so-called 'creativity'". Haddon and Lytton (1968) pointing out that there is no clear evidence for the assumption that high scorers on tests of 'divergent thinking' will be particularly fertile in creative original production in their own life situation also chose to "avoid undue claims" by using the term 'divergent thinking', rather than 'creativity' throughout. This is the point of view also taken in this study.

While, however, it is considered unwarranted to apply the comprehensive label of 'creativity' to divergent thinking tests it is suggested that they are likely, as Burt (1962) observed "to elicit supplementary activities that are rarely tapped by the usual brands of intelligence test". It is also suggested that divergent thinking abilities are likely to be components of value in creative activity, and their concurrent validity in this respect will be investigated later. Their construct validity in relation to a process of creative thinking will be discussed in the next chapter.

Although it will be impossible to avoid using the term 'creativity' when referring to other researches they will be interpreted in terms of divergent thinking whenever possible. The use of the word 'divergent' instead of 'creative' to describe the thinking implicit in divergent tests does not remove the necessity to provide evidence regarding its relationship to other variables, including practical performance, but it removes the intrinsic assumption implicit in the label 'creativity' that a divergent mode of thinking is necessarily directly related to creative work in an aesthetic or professional sense.

CHAPTER THREE

THEORETICAL APPROACHES TO CREATIVE THINKING

It has been maintained in the last chapter that the distinctive feature of divergent thinking tests, that subjects produce a large number of different and unusual responses to some stimulus situation, is not sufficient to warrant the tests being labelled as measuring creativity. There is however a great deal of theoretical and anecdotal evidence built up over centuries that gives credence to the belief that such an ability is an ingredient in the thinking processes which have been described as characteristic of those creative individuals whose creativity is beyond doubt, and in the thinking processes which others postulate are essential for cognitive excellence.

It is in this sense that the writer suggests that divergent thinking tests are of relevance for creative activity. They will not provide a sufficient criterion for predicting real-life creative eminence, but they may well provide a means of assessing some of the necessary ingredients.

The construct validity, or as Cattell (1964) would prefer, the 'concept' validity of divergent thinking tests depends a great deal on the theoretical evidence that can be built up or 'constructed' to substantiate the concept. A certain amount of evidence has already been presented to give plausibility to a divergent aspect of creative thinking and to sustain the discussion up to this point. It is now intended to look more closely at some of the relevant background and theoretical approaches to divergent factors in the study of creative thinking abilities.

3.1 DIVERGENT THINKING AND GUILFORD'S STRUCTURE-OF-INTELLECT THEORY

In his 1950 Address, Guilford attempted to explain creativity in terms of a "factorial conception of personality" in which all individuals possess patterns of primary abilities which govern their capacity for creative thinking. While he notes that creative production in everyday life is undoubtedly dependent upon primary traits other than abilities, such as motivational and temperamental variables, he restricts his analysis to cognitive factors. Having noted that any list of creative thinking abilities would include reasoning factors he emphasises his hypotheses of "other possible thinking factors that are more obviously creative in character". These, he suggests, would include sensitivity to problems, ideational fluency, ideational novelty, synthesizing ability, analysing ability, reorganising or redefining ability, span of ideational structure, and evaluating ability. In addition he suggested appropriate tests which, largely novel at the time, have now become very familiar, such as suggesting improvements to a common article, naming as many objects as possible with a given property, or giving consequences to some hypothetical occurrence such as 'no need to eat'.

Although there are thousands of observable traits Guilford suggests that many are interrelated and that by intercorrelation procedures it should be possible to determine the threads of consistency that run through the various categories and reduce the number of variables. He therefore proposed that a fruitful approach to the domain of creativity would be through a complete application of factor analysis, beginning with hypotheses and tests concerning the primary abilities and their properties.

Guilford's explorations bore fruit in the form of his structure of intellect model (Guilford, 1956) which was then developed into the well-established three dimensional form shown below (Guilford, 1959, 1968).

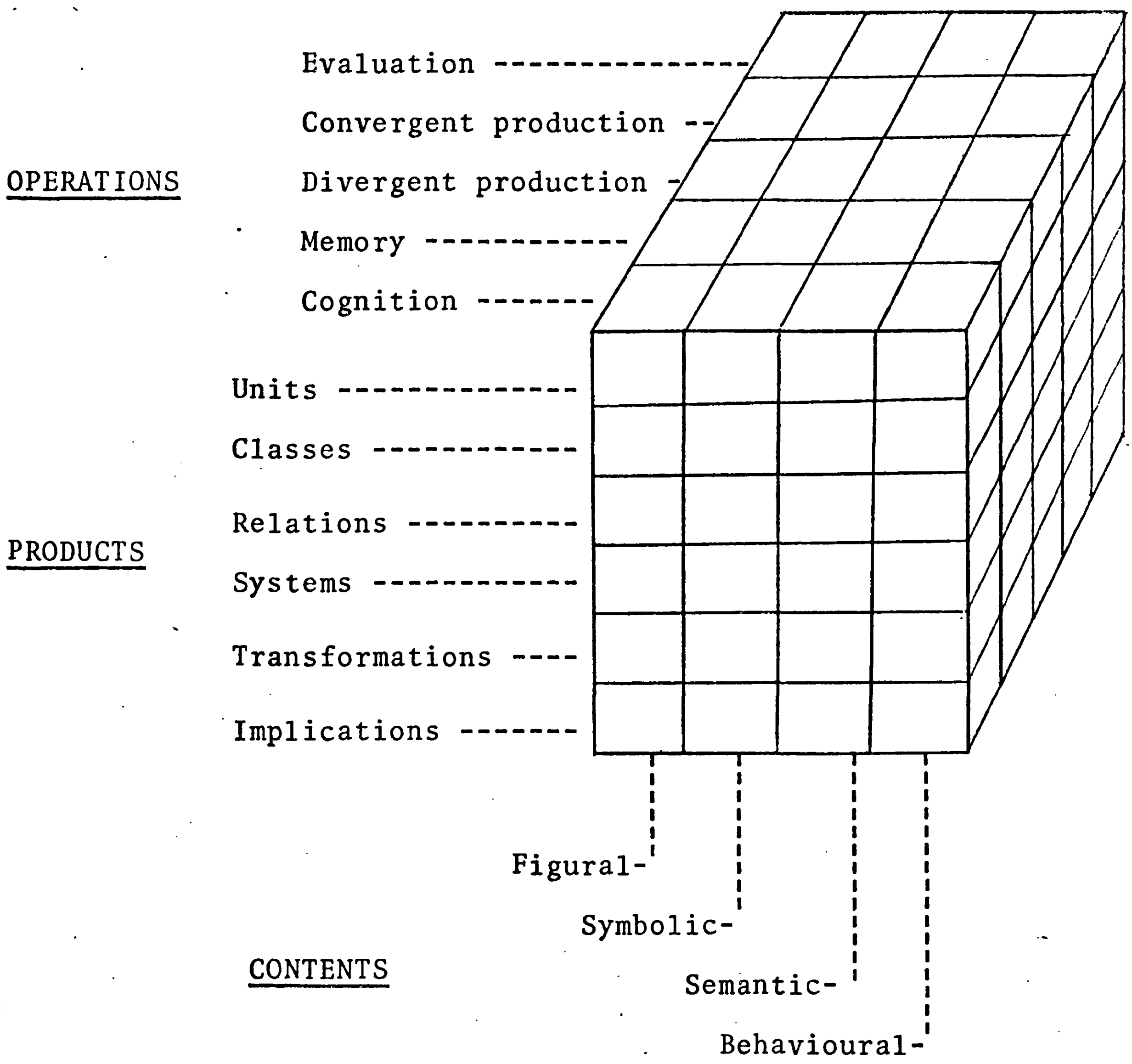


Fig.1. Guilford's Structure of Intellect Model

Each intellectual component or "cell" (or 'factor') in the model represents a unique ability that is needed to do well in a certain class of tasks or tests. Guilford (1959, 1968) maintains that each ability is sufficiently distinct to be detected by factor analysis, but that they also resemble each other in certain ways which give rise to three classifications, according to the type of 'operation' performed, the kind of material or 'content' involved, and the type of 'product' produced.

(i) <u>Operations</u>

There are five kinds of intellectual 'operations', cognition, memory, convergent thinking, divergent thinking, and evaluation. Cognition means discovery or rediscovery or recognition. Memory means retention of what is cognized. Two kinds of productive-thinking operations generate new information from known information and remembered information. In divergent-thinking operations we think in different directions, sometimes searching, sometimes seeking variety. In convergent thinking the information leads to one right answer or to a recognized best or conventional answer. In evaluation we reach decisions as to goodness, correctness, suitability, or adequacy of what we know.

(ii) <u>Content</u>

The four kinds of content are described as figural, symbolic and semantic, and behavioural. Figural content is concrete material such as is perceived through the senses. Symbolic content is composed of letters, digits, and other conventional signs, usually organized in general systems, such as the alphabet or the number system. Semantic content is formed of verbal meanings or ideas. The fourth category of behavioural content is largely theoretical but is included to cover the possible development of the concept of 'social intelligence' and its social or behavioural content such as thoughts, feelings, and perceptions of others.

(iii) Products

When a certain operation is applied to a certain kind of content six types of product may result: units, classes, relations, systems, transformations and implications. Convergent production of symbolic relations, for example, would result from an analogies test item that called for completion of the following:-

pots, stop   bard, drab   rats ..?..

Each cell in the 5 × 4 × 6 model can be designated by its position relative to the three 'axes' of operation, content and product respectively, and represents an unique combination of these properties. The divergent thinking category constitutes one layer of twenty-four different cells. This category is our chief concern and is now looked at more closely.

3.11 The Category of Divergent Thinking

Guilford (1970) defines divergent thinking as *"generation of information from given information, where the emphasis is upon variety and quantity of output from the same source; a search for logical alternatives"*. His interpretation of 'logical' in this context is in the weak sense of 'relevant' alternatives, and divergent thinking is contrasted with the accompanying definition of convergent thinking as the "generation of information from given information, where the needed information is fully determined by the given information; a search for logical imperatives".

In divergent thinking one's search for ideas is manifest in the form of *fluency* in the production of information, *novelty*, *flexibility* and *elaboration*. (Guilford, 1965, 1968.) According to Guilford's original (1950) hypotheses the more creative individual would think with greater fluency, with more flexibility and with greater originality than his less creative peers. The quality of elaboration was added later (Berger, Guilford and Christensen, 1957) as factorial investigations revealed that in tests calling for a variety of answers the degree of extra detail supplied with the responses constituted a distinguishable ability.

The unique feature of divergent production is that a variety of ideas are produced. Tests of this ability thus require the subject to think of a number of answers in reply to some initial stimulus. He might, for example, be required to name all the things he can think of that are round, or give a list of titles for a picture or story. His answers are then assessed in terms of criteria such as the number of ideas he can think of, their variety, and their originality.

While fluency is assessed in terms of the total number of appropriate ideas a subject can think of, usually in a given time, flexibility is concerned with how well he changes the direction of his thinking in some way. It can be a change in the meaning, interpretation, or use of something, a change in understanding of the task, a change of strategy in doing the task, or a change in one's interpretation of the goal. "How often," Guilford (1970) asks, "do we keep trying to solve the wrong problem?" Solving a problem may very well depend upon revisions in one's conception of the problem, and changing one's conception is a matter of intellectual flexibility. A parallel with the Gestalt conceptions of 'restructuring' and 'functional fixedness' will appear later.

Originality in the context of divergent tasks is interpreted by Guilford (1962) as the production of 'unusual, far-fetched, remote, or clever responses'. It is 'originality' within the context of the person's own development rather than in an absolute sense of being previously unknown to mankind that Guilford (1965) maintains is important for assessing creative abilities. Neither of these criteria are possible to apply in practice however, as we cannot be sure whether an idea has existed before or is new to the person concerned. Empirically therefore, Guilford resorts to the concept of originality as statistical infrequency within a population which is culturally relatively homogeneous.

Although Guilford (1958, 1968) emphasises that it is in the divergent thinking category in his structure of intellect model that we find the abilities most significant for creative thinking and invention, he also points out that creativity is not a *single* variable, but "is many things and takes many forms". Each of the categories in his SI model contain abilities which he claims are relevant to creative production, including the twenty-four factors in the divergent thinking category itself. Twenty-three of the latter he has confirmed by factor analysis, though this is very different from demonstrating any practical usefulness for this number of factors, - which even Guilford himself notes are only more or less independent (Guilford, 1970).

The divergent thinking abilities assessed by the tests to be used in the present study are located in Guilford's theoretical model as follows:

Circles Test (Guilford, 1967; Guilford and Hoepfner, 1966; Torrance, 1962, 1974)

'Draw on the circles provided to make them into as many different objects as possible'.

This test involves divergent production of figural material. Guilford (1967) notes that when scored for fluency it provides a good measure of divergent figural units (DFU), though he gives no information on the factor composition of the flexibility and originality scores for this specific test. Flexibility scores seem however to imply a fairly obvious loading for divergent production of figural classes (DFC), though the originality score could imply the production of both systems and transformations. The former is concerned with a 'reorganisation of parts into new wholes', while the second 'reinterprets lines or elements in a different way'. Discussing an alternative figural test (Making Objects), which involves using predetermined parts to make different objects, Guilford (1967) notes that what were thought in principle would be transformations turned out in practice to be systems. The distinction, he suggests, depends on whether the elements are extracted from one figure and then reinterpreted, or whether they are originally seen to exist as separate entities. In the Circles Test this distinction is not easy to make for the originality score could arise from either activity. Guilford's answer would be that the originality score is not unidimensional enough for this test to be used as a specific measure of the cell for Divergent Production of figural transformations in his model. It is this procedure of being able to dispense with tests which do not fit exactly into one's theoretical model which indicates some of the arbitrary nature of the factor analyst's approach.

<u>Consequences Test</u> (Guilford, 1950, 1951, 1967; Torrance, 1962, 1974)

'What changes would occur if we had no need to eat?'

Guilford (1967) reports that this test is concerned with divergent production of semantic material, but that it is again not entirely unidimensional. It gives rise to DMU (divergent production of semantic units) if scored simply for number of responses, but this score also has secondary loadings on the cognition factor for semantic implications. When scored for unusual consequences it provides a measure of divergent production of semantic transformations for which Guilford reserves the word originality.

<u>Uses Test</u> (Guilford, 1950, 1951, 1967; Torrance, 1962, 1974)

'How many uses can you think of for a newspaper?'

This test, involving divergent production of semantic material, gives rise to units when scored for fluency, to classes when scored for flexibility, and to transformations when scored for remote responses. In order to justify the 'originality' label Guilford maintains that it should be clear that there is a shift of meaning to something novel, unusual or clever. The flexibility score, described by Guilford as "adaptive flexibility with semantic material", has some overlap with scoring for transformations, though Guilford claims it is still a good measure of divergent production of semantic classes (DMC). It is likely that an originality score would reflect both a shift of meaning to determine the unusual use and a shift of classes, and would not be a distinct measure of DMT. Guilford (1959, 1973) acknowledges that there is a possibility that "what we have called originality is actually a case of adaptive flexibility when dealing with verbally meaningful material". In this test the quality of output, or more accurately, the quantity of high-quality responses is therefore likely to be an indication of both originality and flexibility in production.

3.12 Discussion

The details of how test scores are interpreted in Guilford's model have been necessary to illustrate the precision of the model and Guilford's suggestion that "the major implication for the assessment of intelligence is that to know an individual's intellectual resources thoroughly we shall need a surprisingly large number of scores". (Guilford, 1959, 1968.) It has also been the writer's contention, expressed in the last chapter, that investigations of creativity should acknowledge a variety of intellectual abilities including specific rather than general skills involved in creative thinking. Guilford's model however, while being a major contribution to any systematic analysis of intellectual abilities, takes this position to something of an extreme. In the selection of 'high-aptitude personnel', the area in which Guilford's research originated, a multiple-score approach to the measurement of abilities may be useful for selecting people with special skills for very specific tasks, but for general use the 120 different intellectual factors are likely to be too highly specific.

Guilford's theory of the intellect however has been the stimulus and the foundation for a large proportion of the research into creative abilities with the focus on his category of divergent thinking. Although different researchers, such as Wallach and Kogan (1966) and Torrance (1974) base their measures of divergent thinking on modifications of Guilford's principles an understanding of Guilford's theory is essential if subsequent modifications are to be seen in perspective.

One of the paradoxical ways in which Guilford's theory and test construction has been developed is remarked on by Butcher (1972). Whereas Guilford has been the main antagonist of an unitary concept of general intelligence or general convergent abilities, and has gone to the opposite extreme in fragmenting abilities, Butcher points out that some writers "have, on rather weak evidence derived from the use of Guilford's test, postulated a trait of *general* creativity or *general* divergent thinking" (original emphasis). Some of the paradox is explained by the very impracticality of assessing the individual abilities covered by Guilford's theory.

Investigators who have followed Torrance (1962, 1974), commonly adopt only the three divergent scores of fluency, flexibility and originality, summing the scores for fluency etc., from each of the sub-tests in their divergent test batteries. Others, including the writer (Richards and Bolton, 1971; Nuttall, 1972) have suggested that these abilities are more task specific, especially between verbal and figural tests, and maintain individual scores for each test. At the other extreme it is not uncommon to find users obtaining a single composite divergent thinking or 'creativity' score. Commenting on this practice, Torrance (1974) observes that while the latter does appear to give quite a stable index of a person's 'creative energy' he recommends users to base their interpretations on the separate scores for fluency, flexibility, originality and, in some cases, elaboration. It is otherwise possible for individuals who obtain high scores for fluency by producing a high number of obvious and trivial responses, to be placed in the same category as those who give a smaller number of responses but gain a similar overall score due to better flexibility or originality.

The relationship between fluency, flexibility and originality scores will be looked at empirically later in this study, and the relationship between convergent and divergent scores should cast some further light on the question of whether the categories overlap or are relatively distinct.

McNemar (1964) argues that provided there is not a drastic curtailment of range a general factor is almost invariably present in tests of cognitive abilities. Though not as productive as Guilford's paper fourteen years earlier, McNemar's (1964) Presidential Address to the American Psychological Association was a timely influence in calling for a better balance in the way creative abilities should be viewed. Highly critical of some of the "half-blind logic" that he suggested characterises some 'creativity' research, particularly that of Getzels and Jackson (1962), he suggested that the concept of general intelligence "still has a rightful place in the ... practical affairs of man". Guilford's structure of intellect with 120 different factors may very well, be suggests, "lead the British, and some of the rest of us, to regard our fractionization and fragmentation of ability, into more and more factors of less and less importance, as indicative of 'scatterbrainedness'".

Burt (1962), Vernon (1964) and Eysenck (1967), though less stringent in their criticisms than McNemar, make similar objections to Guilford's multiple-factor theory and favour a hierarchical organisation of abilities. Eysenck points to the fact that utilizing appropriate testing batteries, there is a possibility of infinite sub-division inherent in factor analytical approaches to studies of the intellect, and that evidence is lacking that further and further sub-factors add anything either to the experimental analysis of intellectual functioning or the practical aim of forecasting success in intellectual pursuits. What is worse, he suggests, is the fact that models such as Guilford's fail to reproduce the essentially hierarchical nature of the data, positive correlations being present among all the relevant tests. Positive correlations among cognitive tests on the other hand can be partly explained by the written format and time limits of most tests, implying some common variance due to test-taking skills.

Cronbach (1970) in a more favourable review of Guilford's theory accepts that while the elements in Guilford's "fine-grain analysis" are not likely to be practically useful in themselves some hierarchical reassembly into groups of abilities may serve as coherent predictors. To determine which regroupings can account for enough variance to be useful will, Cronbach suggests, need a whole generation of validation studies, and the divergent thinking category deserves particular attention.

Having established the framework within which Guilford's divergent thinking abilities have been postulated, the remainder of this chapter will consider some of the other theories that may provide support for believing that these abilities, involving the fluent, varied and unusual production of ideas have relevance for creative activity. A large number of these stem from considerations of creative thinking as a problem solving ability.

3.2 CREATIVE THINKING AND PROBLEM SOLVING

To the psychologist, Vernon (1973) suggests, creative thinking is merely one of the many kinds of thinking which range from autistic fantasy on one hand to logical reasoning on the other. Indeed as he mentions it seems to some extent to partake of both extremes. Vinacke (1952) for example more specifically suggests that "Creative activity can best be understood if it is defined as a combination of problem solving and imagination".

The early history of the concern with thinking processes that are involved in creativity (e.g. Wallas, 1926; Patrick, 1938) tended to make use of concepts drawn mainly from the study of problem solving. Wallach (1970) however is doubtful whether the use of models drawn from general problem solving can succeed in describing thinking processes that differ from the more mundane ones. Most problem solving skills he suggests have been effectively described within the concept of general intelligence and the abilities so described demonstrated, in the sense that individuals can be reliably ordered in terms of their overall competence at coping with the problems offered them by their environment.

In widening the scope of problem solving to include "problems offered them by their environment" Wallach's observation becomes questionable. As Guilford (1958, 1968) argues, the environment would seem to contain far more problem solving activity than of a purely logical reasoning type. In arguing for thinking abilities over and above those involved in logical reasoning however Wallach is drawing a parallel with Guilford's convergent reasoning ability and divergent abilities, though he conceives of the latter as outside rather than within a broad concept of intelligence.

With regard to the creative process however there is little separating the views of Wallach and Guilford that some 'extra ability' is needed to elevate 'routine' problem solving to the realm of creative activity. Gutman (1967) sums up this relationship by suggesting that while creative activity is more than problem solving, the latter is certainly a part of it.

How great a part, and to what extent it involves thinking abilities other than formal reasoning, however, depends on one's interpretation of problem solving. To some writers the activity of problem solving is used to define what is meant by thinking. "Thinking may be provisionally defined as what occurs in experience when an organism, human or animal, meets, recognises and solves a problem" (Humphrey, 1951). Dewey (1910, 1933) also defines thinking as what happens when we solve a problem though he also formulated five stages in problem solving which distinguishes the thinking involved from a more commonplace definition of thinking as "any course or train of ideas" (Drever, 1952).

Dewey's analysis of problem solving consisted of five stages:

1. Recognition of a problem: Occurring in some disturbance of perplexity, doubt, confusion or recognition of a need.
2. Analysis of the problem: A period of searching, enquiring, and assembling of material bearing on the problem.
3. Suggestion of possible solutions: Hypotheses for solution are made.
4. Testing of the consequences: The possible solutions are elaborated and tested.
5. Judgement of the selected solution: The solutions resulting from Stage 4 are evaluated by overt or imaginative action.

By 'imagination' in Stage 5, Dewey means a mental process of evaluation as opposed to an overtly demonstrable one. He deliberately dismisses the more 'day-dreaming' aspects of imagination as unworthy of the name of 'thought'. Similarly Morse and Wingo (1962) state that "by thinking we do not mean the stream of consciousness, nor do we mean ... day dreaming and woolgathering ... We mean the deliberate application of reflection to the solution of a problem".

This interpretation of problem solving could be interpreted largely in terms of convergent thinking abilities though it is possible in Dewey's first two stages to imagine a mental activity in which as in divergent thinking "considerable searching about is done" (Guilford, 1959, 1973). Dewey's stages are in fact reflected in those stages of creative thinking proposed by Wallas (1926) and Rossman (1931). While Rossman like Dewey includes the solution stage or the "birth of the new invention" as occurring after predominantly conscious stages of logical reasoning, Wallas includes an 'incubation' stage in a pattern of 'preparation', 'incubation', 'illumination' and 'verification'.

It is in the 'incubation' stage with its associations of unconscious thought and freely flowing imagination that some writers see the distinction between the steps involved in problem solving and those in creative production.

Whether one conceives of creative thinking as problem solving plus imagination or whether one incorporates an 'insightful' stage into one's concept of problem solving and equates problem solving with creative thinking, both interpretations point to some activity which is over and above that of logical reasoning. A large number of psychological theories have attempted to illuminate the nature of this 'extra something'. Guilford with a multifactorial view of intellect emphasises the role of divergent abilities. Gestalt psychology emphasises that creative thinking entails the restructuring of a problem situation so as to see the elements in a new light and achieve 'insight'. Association theories emphasise the productive combination of freely flowing ideas. Psychoanalytical approaches underline the importance of unconscious conflicts that can either prevent the release of unconscious associations and block the creative process, or else spur it on. Cognitive psychologists emphasise the role of 'direction' and 'cognitive style' in productive thinking.

These differing approaches do not yield independent theories of creative thinking, on the contrary an idealistic picture of an individual engaging in creative activity would draw on something from each of them. It is convenient however to highlight some of these contributions as they relate to divergent thinking abilities.

3.3 UNCONSCIOUS DRIVES AND MOTIVES IN CREATIVE THINKING

Without being drawn into a discussion of creative thinking from an entirely psychoanalytical standpoint, there is a good deal of evidence to support the orthodox Freudian belief (1949) that the energy generated by the unconscious is the motivating force of both the creative person and the neurotic. Vernon (1973) notes that there is an underlying theme in the history of genius that gives some credence to Dryden's line that 'Great wits are sure to madness near allied', Kretschmer (1931) for example being convinced of a psychopathetic element associated with a high degree of talent. Behrens (1974) relating a large number of instances of creative invention from a variety of fields, illustrates the act of genius in going beyond the information given and establishing new closures and acts of imagination in forming new connections and associations. While Leonardo da Vinci, however, in his genius could relate his imagining of battles and miniature landscapes in the stains of rain-soaked walls to the reality of his art, the imaginings of others, in Rosarch's inkblots for example, are often taken as an index of disorder.

Divergent thinking tests do not have quite the same scope as Rorschach as a projective technique but they do assess the subject's ability to produce ideas with only a minimum of restraint. It is also possible, as Hudson (1966, 1968) shows, to analyse the responses to divergent tests in a clinical fashion in terms of characteristics of the subjects such as humour, violence, hostility and authoritarianism, and groups identified by the more conventional measures of fluency and flexibility tend to have common factors in terms of such characteristics. Behrens (1974) points out a number of links between the creative production of ideas which are 'ordered' and recognised as creative, and those which remain 'disordered'. In the notes of William Blake two maxims appear, "Wise men see outlines and therefore they draw them", and also "Mad men see outlines and therefore they draw them". Shakespeare in *A Midsummer Night's Dream* writes that "The lunatic, the lover, and the poet, Are of imagination all compact", all give names and habitations to "the forms of things unknown". Blake, Ezra Pound, and Shakespeare, Behrens suggests were men of uncommon genius mixed with a touch of madness, while on the other end of the scale, we find the idiot savant - overwhelmed by madness, confused by a touch of genius. Both lunatic and genius deny established groupings of ideas but the schizophrenic cannot break out of his reordering of the world which is a compulsive revelation, inadequate for social survival. The grouping and association 'errors' of the schizophrenic are consequently referred to as 'disorders of association' while those of the genius are recognised as being creatively ordered.

Modern developments of the classical Freudian theory, however, are less prone to couple creativity with the neurotic elements of the unconscious and some in fact emphasise that the ego of a creative person must be well-balanced, flexible and secure if he is to realise his full potential (Anderson, 1959).

Storr (1972), outlines some of the ways in which various types of 'psychopathology' can provide the motive power which 'activates the creater', but suggests that the creative person may well be the person who is more adept than most at utilising the inner tensions and psychopathogies that we all have, and to do so may mean that the creative are less likely to suffer from mental illness than others. Mental illness itself, he suggests, almost certainly interferes with creativity.

Kneller (1966) also emphasises adjustment and flexibility rather than disorder in the creative individual. He suggests that it is in the half-way stage, the 'preconscious', between unconscious influences, linked to repressed conflicts and impulses on one hand, and the conscious, which is conventional and reality orientated on the other, that one's level of creative thinking is determined; and it depends on the degree to which one can operate flexibly in the preconscious; assailed as it is by the opposing forces of reality and the unconscious.

Kubie (1965) puts even more emphasis on the concept of flexibility. "The measure of health," he suggests, "is flexibility, the freedom to learn through experience, the freedom to change with changing internal and external circumstances, to be influenced by reasonable argument or by the appeal to emotions and especially the freedom to cease when sated." This concept of 'flexibility' as a personality variable extends beyond Guilford's cognitive interpretation of flexibility (page 72) though it also incorporates the latter. Both aspects will reoccur later in discussing the Gestalt uses of the term.

There are likely to be unconscious factors influencing people with apparently similar intellectual abilities to reach different levels of creation, but like Kubie psychoanalysts are developing theories that incorporate specific thinking abilities and interaction with the environment. White (1961) suggests that there is a drive of an intellectual nature, 'competence motivation' which stimulates creative exploration and experiment, and Rogers (1954) puts forward a similar drive for 'self-actualisation'.

Rogers also claims that creativity is a sign of optimal adjustment, not maladjustment and emerges when one's personality is uninhibited by neurotic tendencies. This links with the general view that as children grow older in an educational system which does not recognise their autonomy they become more inhibited and anxious and less likely to produce creative work. In order to allay anxieties they adopt a conventional, conformist pattern of behaviour. This is then viewed as being a repressive influence on creativity (Moustakas, 1966).

A low score for originality on a divergent test provides some measure of an individual's conformity (though the ratio of originality score to fluency score might be a better index), and it is tempting to assume that high scores consequently reveal a productive and valuable ability for 'originality'. Without practical validation however it is appropriate to point out that other interpretations are possible. Ausubel (1968) for example suggests that high scores on divergent tests have less to do with creativity than mere glibness, uninhibited self-expression and deficiency of self-criticism.

Hudson (1966) while suggesting that convergers and divergers have different personality associations is careful not to label one as more creative than the other. Eysenck (1967) commends a similar view in preference to associating creative production with one particular mode of thinking. He suggests that the divergent superiority of Hudson's arts specialists might well be explained in terms of their verbal fluency. Though not as sweepingly critical as Ausubel, he does suggest that divergent production is very likely a function of personality particularly extraversion.

Developing out of psychoanalytical formulations of the creative personality, and particularly associating the desire for self-actualisation with emotional health, Golann (1962) designed a questionnaire measure of creative motivation based on an interpretation of the 'creativity motive' in terms of a person's "desire to maximise the experiencing of one's own perceptual, cognitive and expressive potentials". Golann's use of this measure in connection with the concept of flexibility will be discussed later.

Whether or not divergent thinking tests reflect this 'creativity motive' or mere extraversion will be subjected to investigation in the experimental section of this study.

It is perhaps salutory at this stage to be reminded of Rossman's (1931) argument regarding creative production that "the assumption that the subconscious is responsible for the final condition is no answer to the problem". In a search for mechanisms that can operate between conscious and unconscious stages however the associationist tradition has been reawakened as one answer to the creative thinker's need for "a ready flow of ideas" and "the freedom to generate extensive ideational possibilities" (Wallach and Wing, 1969). This is the subject of the next section.

3.4 ASSOCIATION AND CREATIVE PRODUCTION

The principle of association refers to a mechanical process by means of which ideas or feelings are connected to one another in the mind. (Shouksmith, 1970.) The first to formulate the principle explicitly was Aristotle (Vinacke, 1952) and the laws of association maintain that the association of ideas or thought processes are dependent on the ideas occurring contiguously, being similar to or contrasting with one another. Shouksmith (1970) follows the development of the theory by Hobbes and Locke in the seventeenth century through J.S. Mill to Bain, Stout and William James, and concludes that the liking for this kind of approach is still prevalent in psychology today.

Studies of creative individuals have given some support to the role of association with the experiences of Poincaré (1968, 1906) in particular being widely quoted (Hadamard, 1949; Ghiselin, 1955; McKellar, 1957).

Poincaré, explaining his discovery of Fuchsian functions relates how after many unsuccessful attempts at proving their existence he retired to bed one night but could not sleep - "Ideas rose in crowds; I felt them collide until pairs interlocked, so to speak, making a stable combination. By the next morning I had established the existence of a class of Fuchsian functions."

Ghiselin (1955) reveals similar instances in his quotations of accounts of the creative thinking of a variety of eminent men. Einstein talks of "combinatory play" or "associative play" seeming "to be the essential feature in productive thought", Mozart of "ideas flowing abundantly", Dryden of "a confused Mass of Thoughts tumbling over one another in the Dark", and Housman of "springs of ideas".

Wallas (1926) and Hadamard (1949) interpret such evidence as stressing the importance of the preconscious sorting of ideas in the 'incubation' stage. For Hadamard invention takes place in building up numerous combinations in what Francis Galton termed the 'ante-chamber of consciousness' and choosing those which are useful. Rugg (1963) in a similar way emphasises both 'pre-logical' and 'pre-conscious' stages in creative thinking and suggests that "the creative flash of insight takes place in the transliminal, across-the-threshold border between the unconscious and the conscious states".

The study of mental 'images' as elements completely independent from conscious ideas has been largely discredited by behaviourist approaches, though the role of imagination and association of ideas as 'not entirely conscious' activities are retained by a number of psychologists in an attempt to focus attention on those internal activities of creative thinking that contrast with more concrete or reasoning activities. Thompson (1959), for example, considers creative thinking as a process involving a 'switching of gears' as ideas flow between an 'imaginative' and a 'realistic' pole.

McKellar (1957) also approaches the subject through a continuum of ideas flowing from logical reasoning on one hand through creative imagination, dreams and related experiences to the hallucinations of psychosis on the other. He suggests that the formation of creative ideas is then not a matter of mere chance but due to a selection from those among which a solution could reasonably be expected. This gives some weight to Guilford's (1950) hypothesis that being able to generate a large number of ideas (ideational fluency) is one relevant stage in creative thinking.

Drawing on, but also adding to the principle of association, Koestler (1964) argues that creative thinking takes place in relation to some problem when the thinker is able to bring together hitherto separate and habitually incompatible ideas in order to solve the problem. He suggests that the association between previously unrelated ideas, the 'Eureka Act', almost always occurs as a flash of insight and is typically the result of both unconscious and conscious thinking. Ideas which are not commonly associated come together in a process which he terms 'bisociation'. As a typical example he relates, within his framework, the story of Archimedes bringing together the commonplace associations involved in taking a bath with the problem of finding the volume of the King's crown.

It is via the conscious visualisation or verbalisation of previously unconscious ideas that Koestler suggests the final solution is reached. If this is the case then as Osborn (1957) and Guilford (1959, 1973) both claim the more ideas that are expressed verbally the more likely it is that some valuable ideas will emerge. Osborn uses this principle as the basis of his 'brainstorming' technique and Guilford in his divergent measure of ideational fluency. Few real-life problems have unique solutions and in the search for solutions the greater the number of possibilities that are suggested the greater the chance of arriving at a suitable solution. Whether by calling for a free and uncritical flow of associations one is more likely to arrive at insightful solutions than by a directed and controlled approach however is debatable.

Firmly in an associationist tradition Mednick (1962) maintains that creative thinking takes place in forming ideas "into new combinations which either meet specified requirements or are in some way useful". This "associative theory of creative thinking" (Mednick, 1962) is the theory adopted by Wallach and Kogan (1966) as the basis for their investigation of creative thinking.

In order to make an assessment of creative thinking in accordance with his theory, Mednick introduced his 'Remote Associates Test', in which subjects are asked to provide one word which would form an associative bridge between three other words. For example, given 'mouse', 'blue' and 'cottage' the subject would need to find the word 'cheese'. Wallach and Kogan on the other hand suggest that this confirms more to Guilford's conception of convergent thinking than to his general definition of divergent thinking, and they presented their subjects with tasks which asked them to generate as many ideas as possible to meet a certain requirement. They gave their subjects five types of task;

Instances: e.g. 'Name all the round things you can think of'

Alternate Uses: e.g. 'Tell me all the different ways you could use a newspaper'

Similarities: e.g. 'Tell me all the ways in which a potato and a carrot are alike'

Line Meanings:
Pattern Meanings: } 'Tell me all the things you think this drawing could be'

Each of these procedures were scored for fluency and originality, only unique responses qualifying for inclusion in the latter. Wallach and Kogan point out that their procedures derive from the work of Guilford but also underline their differences in allowing the subjects unlimited time, giving the procedures verbally to individual children and emphasising the game-like nature of the tasks.

The tasks, they suggest, should carry certain realistic constraints, but they also observe that the associative process as it appears in the introspections of highly creative individuals seems to possess "some degree of functional autonomy from the observer, rather than being entirely under his control. ... Critical faculties, censors, are to some extent stilled when generation of cognitive elements is to be encouraged". To achieve a balance between these opposing principles they conveyed some constraints in the wording of the task, but endeavoured to make the testing atmosphere as 'evaluation-free' as possible.

Based on this conception of creative thinking Wallach and Kogan gave their procedures to 151 children, 'middle class', aged between 10 and 11 years, and concluded that the results demonstrated a dimension of individual differences which on the one hand possessed generality and pervasiveness and on the other hand was also quite independent of the traditional notion of general intelligence. This dimension, based on the ability to generate unique and plentiful associates, they felt justified in labelling as 'creativity', and they proceeded to find a number of behavioural measures associated with their measure.

Getzels and Jackson (1962) though unable to demonstrate as distinct a dimension of 'creativity' as Wallach and Kogan made similar claims for the importance of recognising creative (divergent) abilities as distinct from intelligence. They rejected however the "associationist theory ... as an explanation of these phenomena". Illustrating their objection in terms of the Alternate Uses test they acknowledge that a child can only use those connections that are part of his repertoire but maintain that the laws of associationism (e.g. frequency, vividness, recency) can not explain why some children give predominantly common uses for a brick while others give uncommon ones.

In their efforts to find a more acceptable explanation Getzels and Jackson also reject the operations of traditional logic and go as far as to suggest that adherence to rules of the latter "help us to understand the *lack*, not the presence, of novelty and originality in the behaviour of our children" (original emphasis). Quoting Hadamard (1949) in support they note that invention often appears to be characterised by "sudden, unlogical, non-experiential, and as it were spontaneous solutions" and that both logic and associationism seem to be relegated to secondary roles behind the primacy of "unconscious or fringe-conscious mechanisms". Referring to Wertheimer (1949, 1961) they suggest the possibility of an explanation of creative thinking in terms of the Gestalt approach. A good deal of Getzels and Jackson's arguments against a theory of creative thinking in either logical or associationist terms are in fact a reflection of Wertheimer's own arguments. Wertheimer sees the creative thinking process as more under the control of the individual than is credited by purely associationist theories and emphasises the dynamic relationship between the individual and the stimulus situation. This typical Gestalt approach is now considered further.

3.5 GESTALT APPROACHES TO CREATIVE THINKING

Most commentators on creativity (e.g. Getzels and Jackson, 1962; Kneller, 1966; Shouksmith, 1970; Vernon, 1973) acknowledge the influence of the Gestalt school in understanding the creative process; and its emphasis on flexibility and the need to view objects and ideas in new ways is of obvious relevance to measures of divergent thinking.

The work of Maier (1930, 1931, 1945), Duncker (1945), and Wertheimer (1949, 1961), is of importance in any attempt at explaining creative thinking and is of particular relevance to the design and interpretation of the practical criterion to be used in the present study. Gestalt psychologists, as Wertheimer (1949, 1961) points out, are concerned with the question of what occurs when thinking really works productively, and what is really going on in such a process - "when one has just had a creative idea, however modest the issue, when one has begun really to grasp an issue, when one has enjoyed a clean productive process of thought".

Maier (1930, 1931) investigating the process of thinking in problem solving pointed out several characteristics of his subjects which have since been widely investigated in terms of 'mental set' or 'rigidity'(e.g. Luchins, 1942; Kellmer Pringle and McKenzie, 1966), or as response bias on 'direction' in thinking (e.g. Bruner, Goodnow and Austin, 1956; Berlyne, 1965). Maier described a number of experiments in which unsuccessful subjects exhibited 'functional fixedness', lack of 'direction' or 'rigidity of set'.

In one experiment subjects were required to join two strings hanging from the ceiling which were too far apart to take hold of simultaneously. One solution was to convert one of the strings into a pendulum using an otherwise superfluous pair of pliers. As Maier (1931, 1968) notes "Presenting the subject with pliers is of no benefit so long as the pliers are seen as pliers. They become useful, however, when they are seen as a weight." 'Functional fixedness' is said to occur when the use of an object in one way inhibits a solution of a problem requiring its use in a different way.

Though at first sight some of the reasoning might have appeared as 'trial and error', Maier maintains that on closer observation and discussion successful solutions were seen to depend on changes in organisation and meaning so as to achieve a new 'gestalt'. Duncker (1965) and Luchins (1942) both confirmed and extended Maier's findings, especially in the effects of initial 'set' and the way in which past experience can give rise to 'interfering habits'. Duncker suggests that the key to achieving a solution lies in the subject's flexibility in restructuring and reformulating the original problem.

Wertheimer (1949, 1961) discussing children's attempts at solving certain problems notes that "sometimes one gets surprisingly fine reactions, which are also evident in the remarks of the subject; sometimes one encounters utter helplessness, surprisingly stupid or blind responses even in intelligent subjects". In productive thinking Wertheimer suggests "items do not remain rigidly identical; and as a matter of fact, precisely their change, their improvement is required. ... Blindness to such a change in meaning often impedes productive processes."

This suggestion is adopted as the basis of one of the experimental criteria, the 'Board Game', to be devised in this study. In order to solve the problem presented the children will have to utilize objects in new ways and appreciate their relatedness to the overall structure - the good gestalt. Wertheimer refers to 'fine, genuine solutions' in which the subject shows a grasp of the overall problem and the ability to manipulate its constituent parts in different ways, as A-responses. Blind, foolish and unsuccessful attempts are classified as B-responses. An extended classification of this type will be used to describe children's attempts at the 'Board Game'.

The psychological issue is basically what decides, in the mind of the pupil, between A and B responses and how does he find the A-solution? Wertheimer offers a number of suggestions:

(i) Productive processes are inhibited by blind habits, school drill, bias or special interests.

(ii) Thinking depends on realising the structural features of a situation and the ability to reorganise, regroup and suggest structural improvements.

(iii) The dynamic requirements of a situation require thinking not in terms of isolated piecemeal ideas but in relation to 'whole-characteristics'.

(iv) Thought processes are not arbitrary but develop constructively in spite of deviations and difficulties.

(v) Productive procedures call for not merely 'piecemeal factual truth' but also for 'structural truth'.

(vi) Productive thinking depends on structural insight, structural mastery and meaningful learning. It needs to 'deal with gaps and trouble-regions' by finding inner structural relations and arranging them accordingly, by dynamically segregating and regrouping parts of the whole, and overall, by looking for structural rather than piecemeal truth.

Like Duncker, Wertheimer therefore emphasises that what matters in thinking is how one applies what one has learned from past experience, whether blindly, in a piecemeal way, or in accordance with the structural requirements of the situation. It is the connection of ideas not simply by chance association but with respect to their inner relatedness, their structural features and reasonable requirements, that Wertheimer sees creative thinking taking place. It is not just the quantity of ideas that is important but the need to distinguish between "sensible thought and senseless combinations".

One can only speculate on how Wertheimer would have regarded divergent thinking tests. To see the 'problem' in a way which is not dictated by blind associations, blind habits and blind experiences is an essential ingredient in Wertheimer's productive thinking and it can be suggested that the diverger is precisely the sort of person who views the problems of divergent thinking tests in such a way, interpreting objects and ideas in unusual ways, yet recognising the 'reasonable requirements' of the situation.

It is interesting on the other hand to note that

Wertheimer also calls for an honest attitude and a sincere evaluation of one's thinking to avoid the "danger of dilettantism, of cheap plausibility". This point is appropriate to scoring divergent tests, especially in judging whether a response, which one may suspect of being less than serious, deserves credit for originality.

A more fundamental point however is that the problems envisaged by Wertheimer involve some hard analytical thinking in "trying to penetrate, to realise and to trace out the main relations between form and task", and divergent tasks are likely to be too trivial to demand this component.

It is suggested however that in providing an initial stimulus situation and the requirement that subjects provide possible 'solutions' to fit some goal such as being 'consequences' or 'uses' or recognisable shapes, divergent tests do require sensible as well as unusual combinations and have some of the ingredients of a problem solving situation. It is precisely because of their emphasis on the 'divergent' rather than the analytical process that divergent tests lack justification as 'creativity' tests in Wertheimer's 'productive' use of the term. They nevertheless incorporate an opportunity for the subject to utilize a number of the free and flexible aspects of thinking that the Gestaltists have signified are important for creative production.

It will be interesting to discover later whether the children's performance on divergent thinking tests has some power in predicting their performance on the practical experiment designed in line with the Gestalt conception of a problem situation.

3.6 STYLES OF THINKING

According to the Gestalt interpretation, whether one's thinking is open, flexible and potentially creative, or whether it is a matter of blind habit, is largely a matter of previous experience. Gestalt psychology however gives us only general indications of what sort of experience is desirable to arrive at a good gestalt, and "more attention needs to be given to the forms of action on the environment through which disturbing patterns can be replaced by satisfying ones" (Berlyne, 1972).

Cognitive psychologists such as Berlyne (1965, 1972), Bruner (1956), and Gagné (1964, 1970) have attempted to find 'mediating processes' to explain the links between learning and thinking and the way in which cognitive styles can direct behaviour. Vinacke (1952) notes that a person's thinking is influenced by his adjustment and motivation and is open to forces which can direct its process 'in keeping with inner needs'. Gagné (1964) discussing behaviour akin to problem solving suggests that after the stimulus is received the individual forms concepts relating to the task in accord with past experience, determines a course or courses of action, makes decisions about possible alternatives and verifies the final result. What directs the approach he suggests is a combination of intellectual and attitudinal factors.

Bruner, Goodnow and Austin (1956) in their 'Study of Thinking' claim that from observations of the way in which people approach problems it is apparent that there are "regularities in decision making". In utilising one of these particular 'styles of thinking' Bruner *et al* conclude that individuals may be concerned with one or other of the following objectives:

(i) to maximise information gained for 'each decision'

(ii) to keep the cognitive strain imposed by the task at a minimum

(iii) to regulate the 'risk' consequent on decision making.

Two of the strategies investigated by Bruner were termed 'focussing' and 'scanning' and it is tempting to draw some parallels with 'convergers' and 'divergers'. The focusser tends to look systematically at the information provided and move on to the next stage logically with minimum risk of being incorrect. The scanner on the other hand tends to jump to conclusions and follow his initial ideas until they achieve success or end in failure. The danger in the latter strategy is that frequent failure can result in withdrawal and disorder, though at the same time the willingness to take risks and to follow one's spontaneous ideas without excessive evaluation may result in new and unexpected solutions. The caution of the focusser may be safer, especially within an education system that puts a premium on the correct and the conventional, but it may lack the spark of originality.

Bruner *et al* also emphasise that past experience appears to be a major factor influencing the way people make their decisions. It can, they suggest, provide a basis for making good decisions but it may also be restrictive and make a person too rigid in his or her approach to problem solving and discovery. A person may, they suggest, be predisposed to flexible, open, productive thinking or be limited to the reproductive utilization of a restricted set of concepts.

Getzels and Jackson come to a similar conclusion about the responses of subjects to their divergent tests. For some, they suggested, the dependence on recent and frequent concrete experiences and of fixed and well-worn logical paths "seems to lead to hackneyed and pedestrian responses" while at the same time some children gave "responses that are new, vital, and, in the limited sense in which we are using the term, creative". The result of Getzels and Jackson's study might have been less widely criticised had they, like Hudson (1966), labelled these contrasting styles 'convergers' and 'divergers' respectively. Instead they labelled the children in the top 20% on divergent thinking measures but below the top 20% in I.Q., as a 'High Creative' group, and children in the top 20% on I.Q. but below the top 20% on divergent thinking as a 'High Intelligence' group. They compared these groups on school achievement, aspects of behaviour and attitudes, and how they were regarded by teachers and parents. The results indicated two different modes of thinking and their subsequent plea for the proper recognition of children with abilities other than conventional intelligence gave impetus to the creativity movement. One has to remind oneself however that each of the two groups contained only about 5% of the total sample, that pupils in the top 20% on both I.Q. and divergent thinking were excluded, and that the whole experiment was conducted with a population of gifted children with a mean I.Q. of 132.

As in the case of Bruner's 'focussers' and 'scanners', contrasting types of cognitive activity can be illuminating but the labels should not be taken as identifying two distinct groups. This applies to other related research such as Witkin's (1965) work on field-dependence and independence, Kagan's (1966) reflection/impulsivity dimensions, Pask's (1969) distinction between 'holists' and 'serialists', Satherly and Brimer's (1971) investigation of 'analytic' and 'synthetic' styles, and Luchins' (1942) discussion of rigidity and flexibility. Perhaps of all the contrasted styles of thinking, with the possible exception of convergers and divergers, the latter has received the most attention from psychologists.

Luchins (1951) maintained that the rigidity/flexibility dimension should be regarded as a cognitive style dependent on a person's perception of the field or structure of a problem, rather than as a general personality trait or some "essence of the organism" operating independently of the field conditions. Other experimenters, however, have utilized the concept of rigidity well beyond the cognitive field to explain social relationships, attitudes and prejudices.

Frenkel-Brunswick (1950) reported findings which related rigidity, or as she termed it 'intolerance of ambiguity' to a person's dislike of conflicting situations. Rigid individuals were reluctant to change set, were unable to learn from changing stimuli and suffered from premature closure. These variables appeared in cognitive, perceptual and social areas and indicate that rigidity is a pervasive part of one's personality structure.

Rigidity is explained by Fisher (1950) as a defence mechanism whereby individuals can avoid the stresses caused by "cognitive dissonance" (Festinger, 1957). Faced with conflicting ideas, some individuals therefore tend to cling to their preconceptions and to what they have learned from authority rather than face up to "cognitive conflict" (Berlyne, 1965). Leach (1964) suggests that some people appear to be more flexible than others not just in their treatment of new ideas but in their whole approach to life.

Gallagher (1964) maintains that implications for children's creative expression are obvious. "Two distinctly different patterns or strategies ... begin in childhood and extend to all aspects of life. One of these patterns is characterised by freedom and striving for expression, the other by caution, and concern for the opinions of others." This suggests that not only are the flexible thinkers receptive of new ideas but they are also more likely to produce them, and divergent thinking tests provide one means of investigating this ability.

Torrance (1974) notes that the flexibility score is intended to represent a person's ability "to produce a variety of kinds of ideas, to shift from one approach to another, or to use a variety of strategies", and the originality score reflects some of the same properties with an emphasis on the number of responses which are "away from the obvious, commonplace, banal or established".

Although a low flexibility score appears to be directly related to a subject's tendency to stick to a narrow range of responses and a rigid pattern of thinking, high flexibility scores may be more valuable if used in conjunction with a criterion of 'quality' of response. Without the latter high flexibility might characterise a person who jumps from one approach to another so frequently that he is unable to concentrate on one line of thinking long enough to make it profitable or original. Details of scoring divergent thinking tests for originality are given later, but it is appropriate to point out that the score is devised so that it also reflects the subject's change of category of response. A subject giving a response which is unique to himself would gain maximum originality marks for that response, but he would not gain additional marks for originality if he repeated the same type of idea. Originality marks for divergent tests therefore reflect flexibility as well as novelty. This substantiates the earlier discussion of the tests in relation to Guilford's S. of I. model (page 76).

Torrance (1974) claims that high originality scorers on divergent tests are those who are able to delay immediate gratification or reduction of tension, in order to get away from the obvious, low quality response. Producing the easy 'conforming' response can therefore be interpreted as a mark of the individual who prefers to avoid any cognitive stress.

Stenhouse (1973) also suggests that the diverger will make unusual as opposed to commonplace responses because of a "withholding of the standard response". This "negativistic-seeming sceptical capacity for doubting the authorities and not accepting the standard techniques" Stenhouse terms a 'postponement' or P-factor. He maintains that education does a great deal to inhibit this factor, producing 'learnedness' rather than 'critical flexibility of intelligence'.

Barron (1953) found that people scoring highly on a rigidity scale tended to prefer the simple to the complex on the Barron-Welsh scale of artistic preference. Barron suggests reasons for this relationship in terms of an underlying variable of originality which he maintains demands the fullest possible utilization of stimuli from the environment. This however, he suggests, is not possible for the rigid individual, who selects out and utilizes stimuli in accordance with his defensive needs. Golann (1962) taking this study further argued that subjects high on "creativity motivation" would prefer stimuli and situations which allow for idiosyncratic ways of dealing with them. Using a revised form of the B-W scale with groups of 11 and 13 years of age he found that children who preferred the ambiguous and complex shapes also showed a preference on his creative motivation scale for the activities that allowed more self-expression and utilization of creative capacity, in contrast to low B-W scorers who preferred more routine activities. High scorers for example showed greater preference than the low scorers for activities such as 'drawing pictures', as opposed to 'colouring in a colouring book', and 'figuring out how to do something yourself' as opposed to 'have the teacher tell you how to do something'. The complete questionnaire (Golann, 1961) is reproduced in the Appendix and will be used later as one of the criteria in assessing the concurrent validity of divergent thinking.

It is suggested that the higher children's divergent thinking ability, particularly in terms of flexibility and originality, the greater will be their creativity motivation.

This chapter began with a discussion of the concept and measurement of divergent thinking, as established within Guilford's theory of the structure of intellect, and has so far been concerned with theories of creative thinking which provide it with some theoretical support. Guilford's model, however, did not arise independently of other attempts at conceptualising and measuring intellectual abilities and it owes a great deal to earlier developments in the field of mental measurement. To conclude this chapter some further evidence for the construct of divergent thinking will be looked at within the tradition of psychometric measurement.

3.7 PSYCHOMETRIC APPROACHES TO CREATIVE THINKING

The origins of creative production have been attributed to sources that vary from 'divine inspiration' to 'devilish influence, depending on the use to which the creative talents were put'. (Anastasi, 1968). The Greeks spoke of a man's 'daemon' (without its connotation of evil) which was supposed to furnish the inspiration for his creative work, and the essence of such views was that creative work of the highest order arises from a process which is absent in the ordinary man. Any study of the creative process with a view to encouraging its development in the rest of the population would consequently, on this basis, be doomed to failure.

Galton's (1869) early work was the beginning of a scientific study of creative ability. He remarked on a range of individual differences and began a quantitative approach to the measurement of his subjects' characteristics, including the "fluency and freedom of their associations". Galton's attempts at measuring psychomotor and intellectual abilities were disseminated by his contact with Cattell, who according to Guilford (1967), was the first person to use the term 'mental test' in psychological literature (Cattell, 1890). With the further developments of Binet in France and Terman in the U.S.A., studies of intellectual functioning, including what Galton termed 'creative power' began to take place within a quantifiable framework of individual differences.

Spearman, for example, in 'The Creative Mind' called for investigations of creativity not in terms of some innate power of genius, but in terms of a continuum of intellectual abilities in which some people would show quantitative superiority.

The development of divergent thinking tests has been within this tradition of psychological measurement in an attempt to assess and understand cognitive aspects of creativity, though for the first half of this century their origins are largely buried by the development of intelligence tests, and by the belief that the 'general intellectual ability' measured by the latter would be of most significance in forecasting future talent and creative achievement.

In general terms measured intelligence has in fact been a surprisingly effective predictor of future achievement. Reviewing such evidence, particularly Terman's 'Genetic Studies of Genius' (Terman, 1925; Terman and Oden, 1947, 1959), Butcher (1968) suggests that the evidence demonstrates overwhelmingly, the value of general intelligence, measured by conventional tests, as the most important psychological variable that can at present be assessed to predict future achievement. In spite of a conviction that other variables such as personality and motivation are of importance he acknowledges (Cattell and Butcher, 1968) that they are so elusive that even in combination their practical predictive efficiency is lower than that of general intelligence.

From another point of view however he notes that the individuals who are the statistical exceptions require detailed study and as Goldberg (1965) points out elsewhere, that Terman's work also demonstrates the rarity and unpredictability of creative and original talent. Although it produced plenty of indication of superior educational and vocational attainment and social adjustment it has not yielded any Darwin, Einstein, Picasso or Henry Ford. Guilford (1950) and Getzels and Jackson (1962) both referring to Terman's studies suggest that had he used criteria of intellectual abilities more specifically related to creative aspects, he might have had more success in selecting a more truly 'gifted' sample. As pointed out by the original researchers however, (Terman and Oden, 1947) the occurrence of individuals with supremely gifted talents is extremely rare and the chances of finding them within any sample of 1000 people is very remote.

When the content and form of administration of intelligence tests are scrutinised however one must have certain doubts about their coverage of creative abilities. Not only have the abilities covered become increasingly tied to logical reasoning, but the format, demanding simply ticks or deletions, prevents the individual from expressing any observations other than those prescribed by the tester.

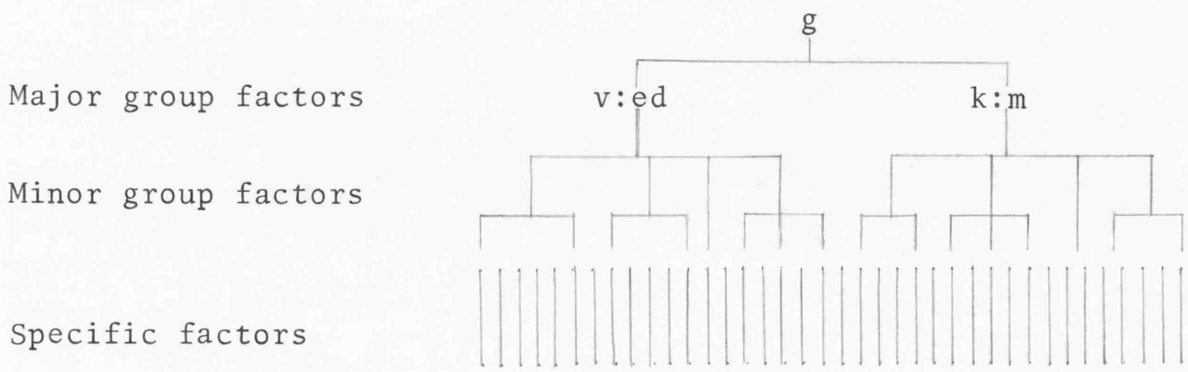
In fairness to some of the early test developers their attempts at measuring individual differences ranged over a variety of abilities and it is in the earlier activities of this century that we find some of the antecedents of divergent tests. Binet and Henri (1896) proposed that ten areas of intellectual functioning be explored: memory, imagery, imagination, attention, comprehension, suggestibility, aesthetic appreciation, moral sentiment, muscular force and force of will and motor skill, and judgement of visual space. For each of these categories they suggested a number of different tests. Their view of intellectual ability was far from a unitary one.

Even when asked to produce a scale for differentiating between different learning abilities within a school situation, Binet's first scale was designed so that the tester could decide how far a child was advanced or retarded intellectually on 30 aspects of mental ability. These covered both psychomotor and cognitive abilities, the latter including tests of comprehension and memory, and three tests which using Guilford's terminology could be classified as divergent thinking. These asked the subject to construct sentences containing three given words, make up their own endings for sentences, and give rhyming words in response to a given word. Under demands for a scale to meet administrative requirements Binet modified and reduced his battery of tests and finally in obvious contradiction to his belief in the complex nature of human intelligence allowed his tests to yield a single score in relation to each child.

Contemporary with Binet's work in France, Terman in America was also investigating the relationship between a variety of mental tests and children's school performance, the latter already beginning to dictate the content of the tests. In one of his early investigations however (Terman, 1906), he chose a battery of tests which would satisfy many advocates of a multi-dimensional view of intelligence. It included what he claimed were measures of inventive and creative imagination, logical processes, mathematical ability, mastery of language, insight (interpretation of fables), learning ability, memory abilities, and motor ability. In order to investigate the power of the tests to discriminate between children's abilities he chose two groups of children whom local teachers rated to be the most and least able respectively. Using teacher ratings to help validate new scales is a reasonable practice, [see for example Entwistle (1968)], but Terman chose only 14 children for his investigation, the 7 brightest and the 7 least able. With these extreme groups it is hardly surprising that most of Terman's tests discriminated between them. The tests of motor ability showed a slight negative relationship however and more significantly (for this survey though not for Terman at the time) the measure of creative and inventive imagination was only slightly higher for the 'bright' group. With hindsight it is easy to suggest that the small relationship between the test of creative imagination and the criterion should have worried Terman more than it apparently did. It is interesting to speculate on how the testing movement might have developed had Terman been less ready to dismiss the creative measure from his battery. Terman might then have been more reluctant to conclude that the similar relationships of the other mental tests suggested that intellectual abilities develop along similar lines and that consequently the measurement of a single intelligence trait was feasible. At the time however the temptation to utilize the apparent success of the majority of the tests would have been difficult to resist, though Kamin (1976) suggests in an historical/political review of the testing movement that the reasons for adopting the I.Q. concept were not altogether scientific. As far as divergent and inventive abilities are concerned Terman's results helped preclude them from most of the early 'intelligence' tests including Terman's modification of Binet's tests into the standardised Stanford-Binet Scales, and they have been so precluded until recently (Warburton, 1970).

These examples of the work of Binet and Terman illustrate two important features. Firstly the initial inclusion in testing batteries of measures of creative ability, and secondly the gradual omission of such measures as attempts to validate and utilize the tests became increasingly tied to the school situation. Furthermore the administrative convenience of using a single score instead of a collection of subscores, and its effectiveness in predicting and discriminating between children's school achievement, laid a foundation for a narrow interpretation of intelligence in terms of a single general ability largely tied to convergent reasoning ability.

As tests of 'general intelligence' began to dominate the mental measurements scene, investigations of thinking more directly related to creative expression and invention tended to take place in terms of specific or group factors of limited scope within an overall model such as that illustrated in Fig.2.



<u>Fig.2</u>. <u>Model of the Hierarchical Group-factor theory of Human Abilities</u> (From Vernon, 1961)

Vernon's familiar hierarchical model of intellectual abilities (Fig.2) developed from Burt (1940) illustrates an alternative theory to that of Guilford. Spearman's 'g' factor of general ability is the most prominent component, in that it accounts for the greatest proportion of difference in abilities, though Vernon also notes that abilities are too varied to be adequately described in terms of a monolithic general intelligence as conceived by Spearman. After the 'g' factor the abilities fall into two major groups, verbal-educational (v:ed) and spatial-perceptual-practical (k:m). These in turn can be broken down into minor group factors typically including scientific, technical, linguistic, physical, musical, or other more specialised abilities. At the bottom of the hierarchy are the specific factors which underlie some particular task but imply nothing about ability at anything else. Vernon (1964) suggests that divergent thinking abilities comprise specific or group factors and that their presence does not conflict with the hierarchical model and the importance of general ability.

Specific factors unrelated to other abilities occur at the bottom of Vernon's hierarchy and although for most general educational and vocational prediction Vernon considers specific abilities to be less influential than abilities higher in the 'tree', he acknowledges that in more homogeneous populations they would be of more relevance in matters of selection and prediction. In the same way Vernon (1969) notes that while there is some divergence of views between British and American factor psychologists regarding theories of the intellect, the American interpretation of intellectual factors as separate primary abilities is a legitimate model when dealing with highly selected or homogeneous groups, where the underlying 'g' factor does not discriminate markedly between individuals.

While McNemar (1964) sees the debate going on for some time as to whether a scheme involving primary abilities plus a de-emphasised 'g' is preferable to one involving an emphasised 'g' plus group factors, Wiseman (1967) suggests that the end-results are largely a matter of emphasis rather than fundamental difference. The essential difference depends on whether in factor analysing a battery of intellectual tests one extracts factors that are distinct or correlated. The British (and Guilford) with a very different emphasis on the initial battery of tests) adopt the former method, the Americans, following Thurstone largely adopt the latter. The end result is either a major general factor followed by group and specific factors or a number of correlated factors of roughly equal weight with an underlying general ability. Thurstone originally suggested seven such factors, spatial abilities, perceptual speed, numerical ability, verbal meaning, memory, verbal fluency, and inductive reasoning.

The number and types of factors extracted however depends on the initial testing battery, and divergent abilities for example, can appear as factors only if they are included in the initial hypotheses and tests of the intellectual abilities to be investigated. Their inclusion depends initially on their theoretical justification, and finally on their practical, predictive validity.

To some extent the justification for divergent abilities is still at the former stage though they have also gained support as questions have been raised about the predictive value of I.Q. alone. MacKinnon (1962) for example reports that studies of eminent men and women showed that, given a minimum level of I.Q. of about 120, there was no relation between I.Q. and outstandingly original work. Similarly Roe (1953) from a study of high level research scientists concluded that "It is then, not essential to have this ability (intelligence) at the highest level in order to become an eminent scientist". Hudson (1966) makes similar conclusions regarding the achievement of able schoolboys.

As in the theoretical discussion earlier there is a definite indication that creative ability is composed of something in addition to reasoning and the type of convergent ability measured by the conventional I.Q. test. Some of this 'extra something' is no doubt due to personality and motivational factors, though it is also likely that explanations of creative ability will be enhanced by a broad conception of intellectual ability including divergent factors.

The researches of Getzels and Jackson (1962) and Wallach and Kogan (1966) brought to the fore the limitations of the traditional I.Q. test and the potential of divergent thinking measures. Both researches claimed to have elicited two distinct types of intellectual ability, the 'intelligent' and the 'creative', and though the latter is an unacceptable label for their 'divergent' dimensions, they were able to demonstrate that the common type of intelligence test fails to sample all or at least a sufficiently broad range of known cognitive abilities (Getzels and Jackson, 1962).

Burt (1962) gives some positive support for the latter conclusion, and notes that "there can be no doubt whatever that these new tests have succeeded in eliciting supplementary activities that are rarely tapped by the usual brands of intelligence test". Defending the general ability concept of intellect however he points out that the conventional intelligence test should not be regarded as an adequate measure of the general ability factor. He suggests, in fact, that "the 'common type of intelligence test' provides a very inadequate measure of that factor", and emphasises, somewhat belatedly, that within the general ability theory there are 'constituent subfactors', which include a group of factors for 'productive' rather than 'reproductive' imagination.

Some early testers did in fact draw attention to these factors, though their practical effect in the development of intellectual tests was negligible.

Hargreaves (1927) gave 151 children tests of imagination, including 'unfinished pictures', 'ink-blots', and story completion. Marking them for 'fluency' he found correlations between all the tests and concluded that "imagination tests, marked for that aspect called 'fluency' had some group factors distinct from 'g'. Spielman and Gaw (1926) arrived at a similar conclusion with tests of 'creative imagination', and although the tests correlated with great intelligence they suggested that there was also a specific factor of imagination to some extent independent of intelligence. Karvé (1929) more forcibly claimed to have "proved the existence of a 'fluency' factor, independent of intelligence, in tests of imagination and association". The latter comprised tests of Controlled Association (write down things made of leather), Free Association, Nouns (give as many as possible beginning with P.T. ... etc.), Unfinished Stories, Ink-Blots, Picture

Completion and Prediction (what might happen if it became unnecessary for people to eat and drink?). These would serve well as a modern test of divergent thinking, yet their existence, and the researcher's conclusions, were largely ignored in practical attempts at assessing intellectual abilities.

It is very likely true as Burt (1962) points out that these abilities can be incorporated within a hierarchical model of the intellect, but it is also true that such abilities were relegated to an insignificant position alongside those cognitive processes assessed by the conventional I.Q. test, and labelled 'general intellectual ability'. Burt suggests that even for predicting scholastic achievement "few educational psychologists would imagine that a single I.Q. based on a 'common type of intelligence test' would yield all the relevant information required", but unfortunately educational psychologists, including Burt, have not always communicated to the community at large the importance of the motivational and special abilities that Burt (1962) maintains should have "considerable weight".

Whatever position divergent abilities finally take up in various theories of the intellect and batteries of intellectual tests, they can be welcomed for the fresh air, which, as Hudson mentions in connection with Getzels and Jackson's research, they let in upon a world of musty, even foetid, expertise. As Heim (1970) has observed the study of abilities appeared to be stagnating with tests being constructed within too narrow a frame of reference. Even if, as Burt claims, the results of divergent thinking studies do not contradict the general ability theory, they have revived interest in factors which have remained well hidden in the conventional approaches to intelligence testing.

3.8 CONCLUSIONS

The theoretical evidence of this chapter has been presented to indicate the associations between divergent thinking and a long line of theoretical approaches to creative thinking. It lends some support to the view expressed in the last chapter that creative thinking is a broad rather than a unitary trait, but also that divergent thinking, with an emphasis on fluent, imaginative, open and flexible thinking is likely to be of an important element in the process of creativity in its broad sense. It has provided grounds for believing that creative ability is unlikely to be encapsuled within the conventional view of intelligence as assessed by the traditional type of convergent I.Q. test, though it has also emphasised that creative thinking, especially as conceived by Guilford in his Structure of Intellect, is not tied exclusively to a divergent category. It is significant that divergent thinking items have now been included in the British Intelligence Scale (Warburton, 1970), one of the subtests including scores for fluency and quality of ideas.

In endeavouring to provide some theoretical associations for the concept of divergent thinking the emphasis in this review has naturally concentrated on the cognitive aspects of ability. On a number of occasions however I have added the rider that real life creative production depends not only on cognitive factors, but also on temperamental and motivational dispositions for creative expression. It is also likely that performance on divergent thinking tests themselves will also reflect the latter qualities, and Golann's (1962) scale of 'creativity motivation' will be used as one criterion in the present study. Heim (1970) suggests that the study of abilities has often distinguished too sharply between the cognitive and affective factors, and Eysenck (1967) maintains that personality variables such as extraversion-introversion and stability-neuroticism should always be included in experimental studies of intellectual functioning, because of their value in mediating predictions and their interaction potential in all types of learning and performance tasks. Some attention will be given to these personality dimensions in the final part of this study in exploring the relationship between divergent, convergent, and attitudinal factors in children's thinking.

It has been suggested that the concept of problem solving seen as the exercise of one's thinking in order to obtain some goal has contributed a great deal to the theory of creative thinking. As Wertheimer points out problem solving is not confined to the schoolroom; problem situations occur throughout life and for a satisfactory outcome, they depend on the thinker getting away from a one-sided inflexible view (recentering), and appreciating changes in meaning of the constituent ideas in accordance with different relationships to the whole structure. In one's "willingness to face issues, to deal with them frankly, honestly, and sincerely", Wertheimer sees problem solving as a "philosophy of life". Education, he maintains, should therefore concentrate not on developing blind, mechanical habits, norms or opinions, but on encouraging children to get to the merit of a situation themselves without depending on the opinions of others, fashion or authority.

While the theoretical evidence does not justify equating divergent with creative thinking, it does give some support to the view that divergent thinking tests are at least pointing in the right direction, and provides some construct validity on which to base the claim that such divergent measures as fluency, flexibility and originality of ideas are of relevance for creative thinking.

The unavoidable issue however is to evaluate their success in demonstrating this hypothesis of relevance. Before proceeding with the practical contribution of the present study the empirical evidence as arrived at by other investigations of their validity and reliability will be reviewed.

CHAPTER FOUR

REVIEW OF INVESTIGATIONS INTO THE VALIDITY AND RELIABILITY OF DIVERGENT THINKING TESTS

4.1 INTRODUCTION

There are a large number of reviews of 'creativity' research of varying depth and range covering studies up to the beginning of the 1970's, such as those of Wallach, 1970; Richards, 1970; Freeman, Butcher and Christie, 1971; Butcher, 1972; Gilchrist, 1972; and Nuttall, 1972. That of Wallach (1970) is a particularly comprehensive review of studies with populations of school children, but with a predictable point of view based on that of Wallach and Kogan (1966).

Whatever their views however, reviewers are consistent in underlining the lack of empirical evidence, referred to earlier in this study, regarding the validity of divergent tests in relation to criteria of creativity. Butcher (1972) for example suggests that far too much has been taken for granted regarding their validity, though he nevertheless accepts that "some connection is plausible" and points to Desing's (1970) study of reliability and validity for a little recent positive evidence.

It is encouraging to observe however that as Vernon (1972) points out, signs of some useful validity are beginning to emerge, and like this study a number of researches are being undertaken to examine the problems of validity and reliability directly. This review will concentrate on recent studies of this type, referring to earlier studies where appropriate.

To begin, an appraisal will be made of the evidence of

Torrance (1974) in comparison with the views of other researchers, followed by a more general review, and concluding with a detailed look at the three validation studies of Bennett (1973), Dewing (1970) and Vernon (1972). Throughout the chapter parallels will be drawn with the present study. In spite of the ceiling placed on their validity by the reliability of divergent tests however, many of the investigations to be reviewed still ignore the latter. Further evidence regarding reliability will therefore be drawn from wider sources when discussing the evidence for scorer reliability, test-retest reliability and long-term stability, in the experimental part of this study.

4.2 EMPIRICAL EVIDENCE FOR THE VALIDITY OF DIVERGENT THINKING TESTS BASED ON THE CLAIMS OF TORRANCE (1974)

Torrance (1966, 1974) in presenting his divergent tests for commercial use has endeavoured to present information on reliability and validity for the information of potential users. Though slender in comparison with the number of studies that have used his tests as criteria of 'creativity' without regard for their validity, there are some studies which provide much needed evidence in these respects.

The greatest number of studies presented by Torrance are in relation to *construct validity*, the divergent tests usually being used on an exploratory basis and revealing interesting relationships with other variables. Most of the researches were conducted by comparing high and low scorers on Torrance's tests, the resulting groups being referred to as the 'High Creative' and 'Low Creative' respectively. Without adopting this prejudicial nomenclature there is a clear indication from most of the studies, as in Hudson's investigations (1966, 1968) in this country, that highly divergent children do tend to exhibit unconventional characteristics both in their responses and in associated personality variables and activities. Most of the studies involving children, however, tended to be with small, selected groups, usually of high I.Q. The divergent tests were, consequently, often being used as an introspective technique with relatively homogeneous groups in terms of intelligence. In these circumstances they appeared to provide a convenient method of discriminating between children in much the same way as the Rorschach Ink Blots. Weisberg and Springer (1961), for example, finding that the high divergers also showed a tendency towards unconventional treatment of the blots.

The half-dozen studies in this context reported by Torrance lend some support to the theoretical evidence reviewed in the last chapter, highly divergent children being found to be less rigid in their attitudes (Fleming and Weintraub, 1962), and able to resist premature closure (Long and Henderson, 1964). On the other hand while two of the studies found evidence of greater humour and playfulness in the young diverger (Torrance, 1962b; Lieberman, 1965), Long *et al* (1965) found the highly original children on figural divergent thinking to be unhappy, lower in self-esteem and alienated from authority.

In studies with older children or young adults groups of highly divergent individuals were again differentiated from lower scorers in a variety of ways. They tended to see themselves as more adventurous, curious and independent and less constrained by authority and others' opinions, than their non-divergent peers (Dauw, 1965), but they also experienced more intense and prolonged stress due to the disparaging views of others, disagreement with parents, and general frustration of their craving for independence (Torrance and Dauw, 1965).

In general the evidence suggests that performance on the tests is related to behaviour characteristics associated with creative thinking, but if it were not for the weight of theoretical backing such as that provided in the last chapter, many of the studies quoted could be dismissed on the grounds of the very limited number of subjects involved, the questionable significance of some of the relationships, the extreme nature of groups selected, and Torrance's lack of evaluation when reporting the results. It is too easy, for example, to seize on, and generalise from, the fact that in one study of a sample of 26 junior high school students, a correlation of 0.52 was found between 'Unusual Questions' and Torrance's Creative Motivation Inventory.

Baird (1972) commenting on Torrance's tests, suggests that in spite of the limited evidence the results seem to be consistent with theories of creative behaviour, though he points to the need for evidence in relation to predictive validity. A very similar assessment is made by Holland (1972) and he concludes that despite gross statistical deficiencies and weak designs in many of the experiments using Torrance's tests, most of the evidence seems generally consistent with the literature of creative behaviour, though lack of external predictive validity is a serious deficiency.

In the latest edition of his tests, Torrance (1974), there are some signs of an increase in *predictive studies*, particularly of a long-term nature. Torrance distinguishes three types of predictive studies, concurrent, short-range (one week to nine months) and long-range.

In the *concurrent studies* with school children, Torrance's criteria consist only of peer nominations of creative behaviour, teacher nominations, and educational achievement. Elsewhere however (Torrance, 1962, 1967), he draws attention to the potential of checklists of creative interests and leisure activities as possible criteria, and it is surprising that no studies using the latter as criteria are reported. Vernon (1964) suggests that research into the home, leisure and educational backgrounds of creative individuals may well yield better measures of creative talent than divergent tests, and some studies involving their use as concurrent criteria will be reviewed later.

Two measures of children's interests and activities will be used in the present study as concurrent criteria. These comprise a 'Creative Leisure Interests Checklist' (Torrance, 1962; Dewing, 1970), and an 'Interests' questionnaire (Barker Lunn, 1970) which includes two groups of creative interests, one 'imaginative' and one 'logical/analytic'. Teacher ratings of the children's creative behaviour are also obtained on a scale constructed by the writer from Torrance (1967).

Defending the use of teacher ratings as criteria of creative talent (with more conviction than that of peer ratings), Torrance (1974) suggests that teachers must be in a favourable position to assess children's creative potentialities, and points to a number of studies giving support to this view, (Yamamoto, 1962; Nelson, 1963). The ratings in Nelson's study, when the teachers gave opinions in a structured manner on scales of behavioural characteristics, are more convincing than those obtained by Yamamoto from general ratings of fluency, flexibility and originality. It has been suggested however (Getzels and Jackson, 1962; Hasan and Butcher, 1966; Cropley, 1967) that teachers find divergent children less desirable than convergent children as pupils, and it is possible that, in giving subjective ratings of creative ability, teachers may not do justice to divergent pupils. Wallach and Kogan (1966) go so far as to suggest that teachers' ratings are "next to worthless" in creativity research. Research findings by Biggs, Fitzgerald and Atkinson (1971) suggest otherwise, however, and lend support to Torrance's view.

Biggs *et al* found that both convergent and divergent children received favourable ratings in terms of classroom behaviour, and concluded that teachers were evaluating divergent children realistically and positively. Although the mean I.Q. in Bigg's sample (108.7 for girls and 104.9 for boys) is higher than in the present study, it is similar in mean age (11.3 years) and background (primary school), and the teacher ratings in this study will be utilized in the belief that they too will be realistic.

As well as the use of teacher ratings as a criterion of creative activity, teacher ratings of the children's personality and desirability in class are also utilized on an exploratory basis in the construct validity investigation.

Studies attempting to validate divergent tests against educational criteria are equivocal both in Torrance's review and in researches using other batteries of divergent tests.

Hasan and Butcher (1966) for example, found correlations of 0.62 and 0.76 between divergent ability and English and Mathematics attainment respectively, while the writer (Richards and Bolton, 1971) found that divergent abilities played only a minor role in mathematics achievement, even in tests designed to assess modern approaches to the subject; and Bennett (1973) reports that, although five divergent thinking measures correlated significantly with conventional English, only one loaded with English on a conventional attainment factor.

The use of educational or academic criteria is complicated by the important role played by intelligence in such achievement, and experiments conducted in groups already selected on the basis of high I.Q. have led to some spurious conclusions about the relative importance of divergent think-abilities and I.Q. [See for example McNemar's (1964) comments on Getzels and Jackson's investigation (1962).] When I.Q. is allowed for, in more representative samples the effect of partialling out I.Q. normally has the effect of reducing the correlations between divergent thinking and academic attainment (Cronbach, 1968; Vernon, 1972), though Bish (1964) reports on the opposite effect.

The effect of divergent ability on educational attainment is an important area of study but a highly complex one, and with differing types of achievement, teaching approaches and ways of learning, Torrance's use of educational attainment as a criterion is not a very satisfactory way of validating divergent tests.

Nine studies, dating from 1961 to 1970, noted as being the major investigations into the *short-range predictive validity* of the Torrance tests are summarised by Torrance (1974). Divergent thinking predicted such creative behaviour of elementary school children as the production of creative ideas about science toys, humour and fantasy, originality in imaginative stories, differential responses to varied curriculum tasks and, in high school students, creative science questions. In the four studies with adult populations, the tests predicted more provocative/divergent as opposed to factual/reproductive questions from six teachers selected as 'high creative' than from six convergent teachers (two studies), teaching success in inner-city schools, and superior performance in productive thinking and creative applications in a Mental Health Course.

The same general criticisms made earlier with regard to the construct validity studies also apply to the above, though as before the overall picture provides some evidence for the predictive validity of the tests.

Results of *long-term validity* studies reported by Torrance are also encouraging though some negative results, and criticisms of Torrance's studies, have been presented by other researchers (Kogan and Pankove, 1974; Crockenberg, 1972).

Of the five long-range prediction studies reported by Torrance (1974), three were by Torrance and his associates and two by other investigators. In that of Witt (1971) Torrance's tests and Witt's 'Favourite Group Games Test' were used to select sixteen black, disadvantaged children from second, third and fourth grades of a ghetto school. After six years, of the twelve who continued with an out-of-school programme designed to provide opportunity for the development of creativity, ten had revealed superior creative talent in the arts and three in science. It provides some support for the efficacy of Torrance's tests in the initial selection procedure but the result is as much an evaluation of the out-of-school programme as it is of Torrance's tests.

Cropley (1972) located 111 students from a sample tested five years earlier on six of Torrance's tests. A composite divergent measure was derived from the test results and compared with data concerning the out-of-school achievements in art, drama, literature and music of the follow-up sample. Significant correlations of 0.52 for boys and 0.42 for girls were found between the criteria and the divergent test predictor. Similar levels of association between predictors and criterion variables are reported in Torrance's own studies (Torrance, Tan and Allman, 1970; Torrance, 1969; Torrance, 1971) but in all these the degree of prediction due to divergent abilities over and above that available from intelligence alone is not determined. Wallach (1970, 1972), however, maintains that the predictive value of Torrance's tests is due largely to their substantial correlations with intelligence.

The earliest of Torrance's prediction studies (Torrance *et al*, 1970) obtained follow-up data after a period of eight years from 114 of an initial sample of 325 student teachers. The students' responses to a questionnaire designed to yield information on 69 different creative teaching behaviours formed a creative teaching index and correlated significantly (r = 0.57) with the original divergent thinking score. Compared with follow-up studies of conventional teacher assessment at college and teaching performance some years later, (e.g. Wiseman and Start, 1965) the predictive validity of Torrance's tests are quite impressive. One would be happier however if the assessments of the teachers' behaviour had been deduced from opinions other than their own.

The two remaining long-term studies reported by Torrance (1969, 1971) involve two groups of children from the same school seven and twelve years after their initial testing in 1959. The first investigation involved 65 subjects, and data, by mailed questionnaire, was received back from 46 of them. The questionnaire asked for information regarding creative achievements, aspirations, and the quality and nature of creative achievements in art, music, literature, science, drama, business and politics. Three measures were constructed from the data: (1) quantity of creative achievements, (2) quality of highest creative achievement and (3) creativeness of future aspirations. These were correlated against the composite divergent thinking measure obtained in 1959, and the resulting validity coefficients were 0.50, 0.46, and 0.51 respectively. The following correlations were obtained for the whole set of predictors in the initial testing battery:

		Criterion Variables (1966)		
		(1) Quantity	(2) Quality	(3) Aspirations
Predictors (1959)	Fluency	0.44	0.39	0.34
	Flexibility	0.44	0.48	0.46
	Originality	0.40	0.43	0.42
	Elaboration	0.37	0.32	0.25
	Composite divergent	0.50	0.46	0.51
	Intelligence	0.22	0.37	0.32

(for 44 d.f. values of $r > 0.29$ are significant, $p < 0.05$)

Table 1 Predictive Validity Coefficients
(from Torrance, 1969)

It is noticeable that the divergent scores, except for the elaboration score, are better predictors than the intelligence score, although there is considerable overlap. It must be noted however that the sample was well above average in intelligence (I.Q. 118) and without this curtailment of range the intelligence effects would no doubt be much greater.

The results of the 12-year follow-up study reported by Torrance (1971) are very similar to the 7-year findings, the data being gathered in an identical manner, and the correlation between the creativity predictors and the criterion variables being 0.51. The study was conducted with a larger population however, of over 200 subjects, and the resulting validity coefficient is even more significant ($p < 0.001$). The greater numbers were only achieved, however, by bringing together the whole high-school population of 392 subjects, with ages ranging from 13 to 19 years. Until such time as this age difference becomes insignificant with regard to future achievement and creative aspiration there is likely to be some age factor in any correlation between divergent thinking performance at school and future creative achievements. This is not allowed for in Torrance's validity coefficient.

As Vernon (1972) points out it is the amount of prediction over and above that which can be forecast by intelligence that is crucial in evaluating divergent tests and Torrance's evidence and that of Cropley (1972) leaves this in considerable doubt. Assuming only a moderate correlation between the divergent tests and intelligence the validity coefficients are high enough to give some positive evidence of predictive validity, but more specific information is needed before any definite conclusions can be made. The effect of intelligence will be looked at closely when validity coefficients are presented later in this study.

4.3 FURTHER STUDIES OF THE VALIDITY OF DIVERGENT TESTS

A study of the predictive effect of divergent thinking and intelligence on creative activities and accomplishments over a two-year period and a seven-year period is reported by Kogan and Pankove (1974) and this study arrives at a very different conclusion to those of Torrance (1969, 1971). As a criterion Kogan and Pankove adopted a biographical inventory of extracurricular activities and accomplishments devised by Wallach and Wing (1969). The latter study provided some concurrent validity for divergent tests in relation to the activities assessed by the inventory, though the experimenters accepted that the interaction between intellective and divergent measures was not established. Kogan and Pankove endeavoured to both extend the validation and ascertain the effect of intelligence. The latter was in fact a combined measure of I.Q. and the power index from the Differential Aptitude Tests.

The testing procedures using the Wallach and Kogan (1966) tests of divergent thinking took place in two different schools at 5th grade (10-11 year-olds), 10th grade, and finally at graduation (12th grade). The predictive validity of the divergent tests (scored for fluency only) and the intellective measures given at both 5th and 10th grades were examined by the degree to which they predicted performance on the activities inventory at the time of graduation. Utilizing a stepwise multiple regression analysis the experimenters found "absolutely no indication in the data of the present study that divergent thinking assessed in the later years of elementary school is prognostic of nonacademic attainment during the high school years". In one school divergent thinking measures obtained at the 10th grade showed a slight predictive effect, significant at the 10% level. Intellective measures were more successful even though the criterion was specifically related to accomplishments in extra-curricular activities such as leadership, art, literature, music, science and dramatic arts. In one school the 5th grade intellective measure accounted for about 45% of the variance in such activities and in the other the 10th grade measure was highly significant ($p < 0.01$). The complete absence of prediction for the 5th grade intellective measure in the larger school however complicated the result considerably.

The results certainly question the predictive validity of the divergent measure, though the researchers' criticism of Torrance's studies, and their own claims, are rather out of proportion to the limited weight of their own study. Out of a total 5th grade population of 46 children in the smaller school only 22 were available for testing at graduation. In the larger school only 46 out of 116 children completed the final testing. In the latter the researchers note that "it was not possible to obtain the cooperation of all of the graduating seniors", and of the 69 who were sent questionnaires (with the offer of $2 compensation for the return of the questionnaire), the 46 who returned them comprised the final sample.

Although both schools were designated as predominantly middle class a considerable proportion of the attrition was due to dropping out of school prior to graduation. Those 46 students remaining and the larger school sample were acknowledged by the experimenters to have a significantly greater level of intellective aptitude than those who were excluded ($t = 3.40$, $p < 0.01$).

The other factor which needs to be recognised when comparing this study with others is the use of fluency only as a measure of divergent thinking. The experimenters note that in a previous study correlations between fluency and originality (uniqueness) had ranged from 0.46 to 0.97 with a median correlation of 0.72, and they considered that this made one of the scores redundant. The degree of overlap, however, could have been revealed more clearly had they included the originality measure in the regression analysis. Whether or not it is valuable to use more than a fluency score is still a matter of debate, but when a number of researchers point to the independent effect of uniqueness on judgements of 'creativity' (e.g. Cattell and Butcher, 1968; Hammaker, Shafto and Trabasso, 1975), an important predictive study such as that of Kogan and Pankove might have included both indices. Wallach and Wing (1969) including both indices in their study found more association between fluency and extra curricular creative activities, than between uniqueness and the latter, but this in itself is not a case for dropping the uniqueness index. Wallach and Wing note that although their index of non-academic accomplishments might not have given sufficient credit for depth of performance, and though they endeavoured to avoid crediting sheer dilettantism, they accept that "as soon as we are talking about level of accomplishment within a given non-academic domain, intelligence becomes more important and sheer fluency of ideas less important". A similar observation could be made regarding the importance of quality of ideas in the divergent test.

Milgram and Milgram (1976) using "a self-report questionnaire of creative activities" owing a good deal to that of Wallach and Wing (1969) found that such creative activity was related to divergent thinking but not to intelligence or school grades. The magnitude of the relationship was not high however and once again the sample was of above average intelligence. No details of the actual I.Q. is given, but the sample was composed of senior high school pupils about to graduate in a "school of high academic standards".

An interesting additional finding was that different scoring procedures designed to credit separately the quality and quantity of creative performances were highly correlated. The procedures were applied to exactly the same data however and some common variance is to be expected from this source. The magnitude of the correlation however was such that the researchers concluded that engaging in a wide variety of creative activities is a necessary condition for unusual and excellent creative attainments. The researchers suggest, somewhat tenuously, that this finding extends to a practical situation the suggestions of Mednick (1962) and Guilford (1950) (as noted in the last chapter), that the production of a large number of ideas is likely to increase the chance of producing original ones.

Though they can provide some evidence of concurrent and predictive validity it can also be argued that measures of creative interests and extra-curricula activities present inadequate criteria of real creative achievement. A recent study (Lang and Ryba, 1976) returns to the creative person as defined by undergraduate specialisation in musical or artistic studies, as opposed to study in 'general arts', in order to investigate creative thinking abilities. One of the creative thinking parameters consisted of Torrance's (1966) tests of creative (divergent) thinking. Inter-scorer reliability exceeded 0.90 for the combined divergent score, though scorer reliabilities for the individual fluency, flexibility and originality scores were not given. The mean age of the students was 21.8 years and the range 18 to 25 years.

The Barron-Welsh Art Scale for simplicity/complexity was also used, the researchers noting its relationship, described in the last chapter, to the cognitive study of personality and to Frenkel-Brunswick's 'tolerance of ambiguity'. After a complex analysis which involved comparisons of means, variance, correlation, factor analysis and discriminant analysis, the researchers concluded that visual artists were superior to the other groups on each of the divergent measures and on the B-W Art Scale, that a general intersensory factor existed amongst creative visual artists, and that certain perceptual factors were shared by the musical and artistic individuals. The former is of most interest to this study, and the researchers suggested that divergent tests and the B-W scale are reliable predictors for identifying the artistically gifted individual. Combining the date from each group, there were strong intercorrelations (0.80, 0.79 and 0.85) between fluency, flexibility and originality, and between each of them and the B-W scale, 0.47, 0.44 and 0.49 respectively. This gives further construct validity to the Golann Creative Motivation Scale (Golann, 1962) which will be used later in this study. As noted in the last chapter Golann's scale was related to creative ability via its relationship to the B-W Art Scale.

As a word of caution, Lang and Ryba conclude their report by noting that, despite the mounting enthusiasm for creativity much spadework remains to be done towards identifying reliable external criteria, and they suggest that investigation of 'intra-intersensory' functions might reveal how a host of different creators:- painters, composers, poets, designers and research scientists - are all able to perceive ratios, permutations and relationships "so as to culminate in the birth of an imaginative Gestalt".

Adopting a similar design to that of Lang and Ryba (1976), Zegas (1976) hypothesised that undergraduates "choosing major study in 'creative' fields (art, music, English) would perform differently on a group of divergent production tests, with each subject doing best on the DP tests whose content area most closely related to his field of study". More specifically, giving divergent tests to a total of 106 undergraduates, to cover figural, symbolic and semantic areas of Guilford's Structure-of-Intellect model, he suggested that they would reflect the different creative abilities of the art, music and English majors respectively. A study of the levels of performance of three 'creative' groups in comparison with a 'general' group supported the hypothesis. There were no sex differences in performance on the tests, though the art majors were all female.

Speed, however, was a major variable in the testing procedures. Ten short divergent tests were given in forty-four minutes, each containing two items. Subjects were asked, for example, to sketch objects from circles in two minutes and similarly from squares in another two minutes. Little information on scoring was provided apart from the statement that "Papers were scored basically on quantity". It is difficult therefore to generalise from Zegas' results though they give some concurrent validity to the fluency aspect of divergent tests given under timed conditions in its ability to differentiate not only between 'creative' and general groups but also between different types of creative group in terms of the separate 'content' categories in Guilford's S. of I. model.

Split-half reliabilities are given by Zegas for each ability within each of the groups of students tested. These are typically high, between 0.82 and 0.93 for total divergent scores in the three content areas, semantic, symbolic, and figural; and between 0.91 and 0.96 for the overall scores.

It has been suggested earlier in this study that divergent thinking ability may aid children's performance on a problem-solving task designed to incorporate flexible and productive aspects of thinking, particularly as suggested by the Gestalt school, and this will be subject to investigation later. Some practical evidence giving some construct validity to this hypothesis is revealed in a study by Goor and Sommerfeld (1976) into the problem-solving processes of creative (divergent) and non-creative students.

From an initial pool of 227 university students the 26 highest and 26 lowest scoring subjects were 'selected' on the basis of two divergent tests, Plot Titles and Consequences.

The tests were scored for originality or 'divergent semantic transformations' by crediting 'clever' titles and 'remote' consequences. The researchers note that they adopted the tests and scoring procedures from Guilford (1967). Apart from providing examples of 'remote' and 'clever' responses however, Guilford (1967) does not specify criteria for evaluating the latter, and Cropley's (1967) suggestion, that a more objective method based on statistical infrequency would be preferable to a subjective assessment, appears to be more widely accepted.

With some doubts about the scoring procedures the investigation nevertheless presents some interesting evidence. The two extreme groups were referred to as 'High creative' and 'Low creative' respectively, though here they are interpreted simply as 'high divergent' and 'low divergent' groups. Both groups were asked to 'talk aloud' as they attempted three problem solving tasks and taped records were kept for the whole experimental period of thirty minutes.

The transcripts were analysed at three second intervals and the type of response at each interval was categorised into one of seven response categories from 'surveying new information', to 'silence'. A Markoff chain analysis showed that different overall patterns of response were statistically differentiable between the high and low divergent groups. When these patterns were looked at more closely three of the response categories differentiated significantly (χ^2 test) across all three tasks. The diverger scored more highly on "generating new information or hypotheses", "developing or working on a hypothesis", and less highly on "silence". Other differences were also found but tended to be specific.

The results give support to the hypothesis that the patterns of thinking of the high and low divergent subjects show discernible differences when faced with a problem solving task, and provide some construct validity for the point of view presented by the writer in the last chapter. It is disconcerting however to find the subjects high on divergent performance being referred to consistently as the 'high creative' subjects and vice versa. We learn for example that the "low creative subject" had significantly higher periods of "silence" than the high creative. One does not have to enumerate the ways in which this can be construed in order to appreciate the dangers. There is no evidence that the high divergent group were any better at solving the problems than the low group. The experimenters hoped in fact that no-one would find a sufficiently plausible solution to make them abandon the task, as this automatically precluded such subjects from their analysis. A total of 12 subjects were in fact thus excluded. As it happened six were in the high divergent group and six in the low group. Apart from the fact that all the subjects were university freshmen, no details of their intellectual ability was provided.

The tasks themselves are familiar to those interested in puzzles or problem-solving; building four equal and equilateral triangles with six matches; the 'cancer' problem from Duncker (1945); and the 'pebbles' problem from De Bono's 'Use of Lateral Thinking' (1967). Only the former has an unique solution. If the subjects remained silent for about 10 seconds the task administrator had a list of "verbal promptings" to be used. Presumably these were used to encourage subjects who quickly proffered 'low level' solutions to the last two tasks to 'think again'.

In contrast to Goor and Sommerfeld's study, the experimental task designed for the present research will be used to assess the quality of the subjects' solutions, but it is encouraging to find some empirical evidence to add to the construct validity for believing that divergent abilities will be a factor in different approaches to the task.

Though not a major area of investigation in this study it was noted in the last chapter that different approaches to creative tasks are also likely to arise from personality variables.

Torrance and Khatena (1970) have developed a personality scale 'What Kind of Person Are You?' incorporating characteristics which are though to distinguish creative from non-creative persons. Halpin and Halpin (1973), and Halpin, Halpin and Torrance (1974) report on its use with undergraduates in exploring the relationship between Torrance's 'creative thinking abilities' and other variables.

In the latter study the personality inventory was significantly correlated with four out of the seven scores from the Torrance tests for males but with only two of the scores for females. While the experimenters note that this data is not impressive if the inventory is interpreted as a measure of criterion related validity, they suggest that it does suggest that the constructs which underline the Torrance tests are significantly related to the constructs of the creative personality as measured by the 'What Kind of Person?' test.

The study by Halpin and Halpin (1973) also fails to report any convincing relationship between Torrance's tests and the personality inventory, performance on both appearing to be affected by the motivation of the subjects. Ward, Kogan and Pankove (1970) on the other hand find that, while motivating conditions may alter the level at which groups perform, individual differences in performance derive from differences in capacity rather than motivation for divergent production.

Turning to more established measures of personality variables, it was suggested in the last chapter that the diverger may exhibit characteristics of the extravert (Eysenck, 1967), and divergers have also been found to be more neurotic than convergers (Hudson, 1968). The latter finding, however, was somewhat blurred by the fact that a minority of neurotic items in the Personality Inventory positively discriminated in the direction of the converger. The results of Smithers and Child (1974) investigating extraversion, neuroticism and divergent thinking in a sample of 306 university undergraduates do not support either view. Identifying the students as divergers (n = 51), convergers (n = 53) and all rounders (n = 202), according to their differential performance on divergent thinking tests ('Uses' and 'Consequences', scored for fluency) and the AH5 test of intelligence, Smithers and Child found no differences in the level of extraversion or neuroticism between the three groups. While this finding was with university students, Hudson's subjects were 15 years of age. Some evidence at the other end of the scale with 11-year-olds will be provided by the present study.

4.4 STUDIES OF THE RELATIONSHIP BETWEEN DIVERGENT THINKING AND INTELLIGENCE

The impetus given to creativity research by the studies of Getzels and Jackson (1962) and Wallach and Kogan (1966) has been acknowledged throughout this study. Most writers would agree with the important educational implications of their findings, particularly regarding the development of abilities other than those covered by the conventional I.Q. test, but their suggestions that creativity, expressed in terms of divergent ability, is a quite distinct and independent entity from intelligence has not been confirmed in various replications and follow-up studies (Marsh, 1964; Thorndike, 1966; Hasan and Butcher, 1966; Ward, 1967; Fee, 1968). At the same time, as in the last two studies mentioned, factor-analytic techniques have demonstrated the presence of some distinct divergent ability loading on factors orthogonal to that of conventional I.Q., though the fact that the divergent variance extends to a number of factors is a clear indication that 'creativity', even in terms of divergent abilities, is not unidimensional (Fee, 1968).

The presence of some distinct 'divergent' variance is crucial to any study of the validity of divergent thinking tests and it is not surprising that the extent of this distinction, both as a study in its own right and in wider studies, continues to be investigated (Anastasi and Shaefer, 1971; Guilford, 1971; Starr and Nicholl, 1975; Guilford, 1975; Richards, R.L., 1976).

In a study of nearly one thousand 15-year-old pupils, Anastasi and Schaefer (1971) compared the responses to the Alternate Uses and Consequences Tests with I.Q. and school grades. They found that divergent tests correlated only as highly with each other as with I.Q., and separate analysis for boys and girls yielded closely similar results. The degree of overlap, they point out, could very likely be modified by a different choice of I.Q. and divergent measures, but they claim that the results indicate that neither dimension exists as a distinct entity. They suggest rather that each dimension includes identifiable traits that, organised in a pattern of relationships, cuts across both domains. This emphasis on divergent thinking as a many-faceted concept is supported by Guilford (1971, 1975, 1976).

Richards R.L. (1976) makes a comparison of selected Guilford and Wallach and Kogan tests, with three general ability measures. Contrary to Wallach and Kogan's recommendations however, their tests were administered in a timed format, of five minutes per item, together with the Guilford tests. The Wallach and Kogan tests were scored for fluency and uniqueness, and the three Guilford tests yielded three scores for fluency, two for flexibility and one for originality. Internal consistency reliabilities were reported for each of the 13 scores. These ranged from a low of 0.45, to 0.81, the lowest coefficients arising from the uniqueness score on the Wallach and Kogan tests. Ten of the coefficients exceeded 0.66, but even so these are lower than commonly reported for internal consistency measures, and Richards used a correction for attenuation to help clarify their relationship with intelligence. As a result she observed that "Results were similar for both Guilford and Wallach-Kogan tests with the Guilford tests showing a slightly higher average relation to general intelligence" (Richards, 1976). There was some slight evidence for convergent and discriminant validity of the divergent measures, the divergent tests correlating more highly between themselves than with I.Q., but having acknowledged the zero-order correlations, of the order of 0.25 to 0.35, between the divergent measures and intelligence, Richards made the surprising conclusion that the "results argue for a relatively simple description of the creative thinking domain, supporting the Wallach-Kogan position of a unitary and independent dimension".

In a reply Guilford (1976) points out the considerable common variance implicit in the relationship between the divergent and intelligence measures, and reiterates his belief in a multi-variate view of both the divergent and convergent dimensions in his Structure-of-Intellect model. He concedes that there may be second-order or third-order factors common to all his divergent production abilities, but when the first-order correlations are of limited size he suggests that the influence of any common underlying ability would be severely restricted.

The discriminant validity of divergent measures in relation to I.Q. and its convergent validity in relation to other divergent tests is often better discussed in terms of the tests factorial validity (e.g. Cropley and Maslany, 1969). Investigating the reliability and factorial validity of the Wallach and Kogan battery of divergent tests, Cropley and Maslany (1969) arrived at results which, though conducted with a highly selected sample in terms of I.Q. (university students), and allowing unlimited time (which varied from 1 hour 15 minutes in the case of one student to 6 hours 30 minutes for another), nevertheless indicated a considerable overlap between the 'creativity' and 'intelligence' domains. The restriction of range and the informal conditions (the students were allowed to smoke, drink coffee and move around the room during the administration), were likely to reduce the I.Q. effects, and the divergent tests did correlate much more strongly between themselves than with I.Q. Cropley's conclusion, that the results suggest that divergent tests measure a stable and internally consistent intellective mode, albeit substantially related to general intelligence, is quite compatible, however, with that of Anastasi and Schaefer (1971), and would also be a more balanced interpretation of the results of Richards R.L. (1976).

Factorial studies, however, do not always clarify relationships between tests though they might raise questions for further study. A study by Plass, Michael and Michael (1974), provides such an example. A factor analysis was performed on the 30 separate measures obtainable from the seven tests in Torrance's 'creativity' battery. Conveniently seven factors were extracted and the researchers conclude that "from the factor matrix presented ... it is apparent that each of the seven rotated factors described each task rather than the hypothesised psychological process for which it was scored".

Their interpretation of the factor matrix however is open to some question. The percentage of variance attributable to each factor is omitted, and while the task specific element is evident, the first factor had substantial loadings from several of the verbal tests. The result of this investigation is a good illustration of the arbitrary nature of the factor analysis technique. Had composite fluency, flexibility, originality and elaboration scores been included to balance the scores arrived at from the same task (nine scores from the same Ask and Guess task), a different pattern would no doubt have arisen.

4.5 EVIDENCE FOR THE TEST-RETEST RELIABILITY OF DIVERGENT THINKING TESTS

The reliability information provided by Cropley and Maslany (1969) relies on an internal consistency (K.R. 20) measurement and applies to the scoring of originality, defined as statistical infrequency. The exact details of Cropley's recommendations for scoring will be discussed later. The K.R. coefficients were 0.67, 0.82, 0.86, 0.87 and 0.85 for the five divergent sub-tests, for the full sample (134 male, 73 female). Although the writer has suggested earlier that internal consistency methods are not the most appropriate for assessing the reliability of divergent tests, Cropley is one of the few investigators who have looked closely at this aspect of divergent tests.

The varied and often unsatisfactory evidence for test-retest reliability has already been outlined in Chapter 1, and little evidence has been presented to aleviate the concern expressed (Cronbach, 1970; Lewis, 1974). Torrance (1974) notes that few test-retest studies of his latest battery of tests have been made. That of Hagender (1967) however reports coefficients which show some improvement over those discussed earlier from Wodtke (1963), and these will be given in more detail later, in comparison with those obtained from the present study. Hagender's results were obtained with a sample of only 28 children after a period of "from one to two weeks", but the verbal test results, in the region 0.79 to 0.87, would be acceptable for most psychological tests. The coefficients for the figural tests ranging from 0.50 to 0.71 were not as satisfactory.

Earlier studies quoted by Torrance (1974) generally have somewhat lower verbal test reliabilities than those given by Hagender, though figural results have been higher. Yamamoto (1962) for example obtaining reliabilities for the Circles Test of 0.76, 0.63 and 0.79 for fluency, flexibility and originality respectively.

The longest time interval between testings reported by Torrance (1974) is that of Dalbec (1966) who, with 43 college students, obtained long-term reliabilities of 0.59, 0.35 and 0.73 for fluency, flexibility and originality, after an interval of three years. Cropley and Clapson (1971) have also reported on long-term test-retest reliabilities.

Out of a group of 320 twelve-year-olds who had taken a variety of tests in 1964, Cropley and Clapson located 110 who were still at school five years later, and investigated the stability of their performance on two divergent tests, 'Consequences' and 'Circles'. The mean I.Q. for the final sample was 119, the population, as Cropley remarks, resembling that of an English Grammar School. Making some correction for the restriction of range, which he also assumed, on the basis of reduced variance, had occurred in the divergent responses, Cropley found reliability coefficients for the whole sample of 0.45 for the Consequences Test, and 0.44 for the Circles Test. Those reported for the boys separately were somewhat higher, 0.58 and 0.48 respectively, while those for the girls were 0.33 and 0.40. Although significant, Cropley remarks that these are rather low, but within the range of test-retest reliabilities commonly reported for subscales of other well accepted mental tests, including those of the WISC.

Haddon and Lytton (1971), retesting a sample of English 15 to 16-year-olds, four years after their initial testing, reported test-retest reliabilities of 0.50, 0.55 and 0.62 for composite non-verbal, verbal and total divergent scores respectively. This investigation will be looked at more closely later when considering the long-term stability of the tests in the present investigation.

Evidence regarding scorer reliability and further details of test-retest reliability and long-term stability will be incorporated in the later chapters devoted to investigating these aspects of reliability.

This chapter will now be concluded with a review of three studies specifically designed to investigate the validity and, in the case of Dewing (1970), the reliability of divergent thinking tests and consequently of particular relevance to this study.

4.6 STUDIES OF VALIDITY AND RELIABILITY BY BENNETT (1973), DEWING (1970), AND VERNON (1972)

Bennett (1971), suggesting that there was an unsatisfactory link between divergent thinking and creativity in one book on the subject, (Lytton, 1971), underlined the pressing need for some validation studies. This was followed up (Bennett, 1973) by a report on one such investigation. 300 ten-year-old children were drawn from eight British primary schools resulting in a sample with slightly above average V.R.Q. (104.8) and standard deviation (17.0). Divergent abilities were assessed from five pairs of semantic divergent thinking tests, taken from Guilford and Hoepfner (1966).

Each of the five abilities tested was provided by a composite score from each pair of tests. No details of the tests are given however, and neither from subsequent mention of the tests, nor from the reference provided, can one deduce precisely which tests were used. Parallel tests were presumably of the nature of 'Uses for a brick' and 'Uses for a newspaper'.

In accord with Guilford's scoring procedure each test was scored for one ability only. Three of the tests gave fluency scores, one an originality score and one a flexibility score. These were analysed separately, and an overall divergent score was also computed by standardising and averaging the five separate ability scores.

The correlations among the divergent tests ranged from 0.36 to 0.74 but they also had significant correlations with V.R.Q. ranging from 0.26 to 0.50. The overall score for semantic divergence correlated 0.54 with verbal reasoning for boys and 0.58 for girls. These are similar to the values obtained by Haddon and Lytton (1968), but the latter also included the results of figural as well as verbal tests in their composite score.

Although the intercorrelations of individual divergent tests are greater than their correlations with verbal reasoning the latter are substantial enough to indicate a good degree of overlap. Analysing the scores in categories, Bennett remarks that only 3% of the sample combined low divergence with high convergence and vice versa. In contrast 18% gained high scores on both ability measures - a category ignored by Getzels and Jackson (1962). Apart from the expressional fluency score however a factor analysis showed some evidence of discriminant validity between divergent and convergent measures. This was reported for both sexes combined, separate factor analyses have been found to be "almost identical".

Having reviled in his introduction the "dearth of objective criteria" on which to base validity studies, Bennett's attempts to provide such criteria are disappointing. He utilizes only an 'imaginative story'. Two teachers "gave impression marks for the use of imagination, good ideas and so on, ensuring that poor spelling and grammar were not penalised". Mark/re-mark and inter-marker correlations both exceeded 0.70. The latter is not a very high level, though impressionistic marking, with no particular guidelines, could not be expected to give a much higher result.

Some evidence of a relationship between the divergent measures and the imaginative story appeared on the third factor (accounting for 17.7% of the variance) of a varimax rotation, and an analysis of mean scores indicated that, given a certain level of V.R.Q., children with high divergence had a better performance on the story than those with a low divergent score. Bennett noted that this gave "evidence for construct validity". This, however, blurs the specific concurrent validity evidence from the 'objective criterion' with that of the general relationship between convergent and divergent measures. Overall the results are not very convincing, Bennett himself concluding somewhat ambiguously that "the evidence concerning the construct validity of semantic divergent abilities is encouraging, but it does not provide any justification for treating such abilities as synonymous with creativity".

Dewing (1970) directed her study more specifically towards criteria of creative performance. As well as an imaginative story, she incorporated peer ratings and teacher ratings of in-school creativity, Torrance's Creative Leisure Interests Checklist, and the Golann Creative Motivation Scale.

Selecting children in the top stream from ten different schools in Perth in Western Australia, Dewing's sample consisted of 394 children aged 11 to 12 years, with a mean I.Q. of 113.9 (s.d. 10.5). Two pairs of divergent tests were used, Alternate Uses for bricks and 'tin cans', and the Circles Test with its parallel form as 'Squares'. Fluency and originality scores were obtained for each, standardised and added to form a total 'creativity' score. No indication is given of the correlation between the verbal and non-verbal pairs of tests and the results might have been more illuminative had they been kept separate. Each test was given a time limit of 12 minutes "by which time a pilot study had shown that response production by most students had ceased".

The measures of creative performance were intercorrelated within each school, with the exception of teacher ratings which were made for only those children rated as being in the most creative 20%. Most of the intercorrelations were positively correlated as one might have expected, but they in no way provided a unidimensional measure of creativity. Some of the intercorrelations were in fact negative. This occurred particularly, in six out of ten cases, between the Torrance checklist and peer ratings and gives some weight to the doubts expressed by Wallach and Kogan (1966) about the value of peer ratings for creativity.

Dewing did not attempt to compare divergent performance with any total creativity index however, but selected the top 20% of the children on the total divergent thinking score and compared their creativity performance with a matched group from the remaining 80%. 61 children in the high divergent group were in the top 20% on at least one creative performance measure as compared with 31 from the control group. The result, further categorised into High, Medium and Low Performance, gave a highly significant result ($\chi^2 = 28.476$, d.f. = 2., $P < 0.001$). A moderate degree of validity in this indicated, creative thinking ability (divèrgent thinking) being significantly related to creative performance as judged by the various measures.

Dewing did not however include any "mediating variables" Eysenck (1967), such as attitude or personality, which might have accounted for some of the difference in performance. It must be remembered that the control group was selected from a population which had been truncated by the removal of those in the top 20% on divergent measures. As Hasan and Butcher (1966) found in replicating Getzels and Jackson's study it is not always possible to find a complete distribution of other abilities, such as I.Q., within the remaining group. Like Bennett (1973), Dewing found no marked sex differences with regard to either divergent thinking or creative performance.

Before reporting the validity results Dewing gave details of the reliability of the divergent thinking measures. Test-retest reliabilities over a six-week interval were reported separately for the Uses and Circles Tests and for originality and fluency. With ten samples ranging from n = 31 to n = 56 the average reliability coefficients for the Uses Test were 0.52 for fluency and 0.39 for originality, with corresponding values of 0.69 and 0.54 for the Circles Test.

No comment was made on the way in which these results might affect the validity conclusions. The evidence itself, however, particularly for the Uses Test is not very encouraging.

Vernon (1972) reports on an investigation into the validity of divergent thinking tests in relation to a large number of variables. This major study involved nearly 400 children of median age 13y 11m in rather a wide range of ages from 12y 4m to 17y 0m; and about 10 hours of test procedures covering over 100 test variables.

To investigate concurrent validity Vernon selected ten variables and summed their scores to form a "scholastic and daily-life criterion of creativity". These consisted of

(i) Essay writing, scored for 'conceptual maturity' and 'imagination and creativity'.

(ii) The rating for imagination and originality on McClelland's Need for Achievement Test.

(iii) The rating for creativity on Witkin's "Drawing of a Man and Woman" Test.

(iv) Sociometric score for choice of 3 persons to work with on a creative project (disguised peer rating).

(v) Ratings by two teachers for curiosity.

(vi) Scores for answers to a questionnaire on leisure-time activities, giving scores for 'artistic-creative' and 'scientific-creative' activities.

In an interesting comparison Vernon reported the following validity coefficients (Table 2) between the criteria and each divergent thinking test, firstly by direct correlation and secondly with the effect of general verbal ability held constant:

	r with Creativity Criterion		Ditto, V.R.Q. held constant	
	Boys	Girls	Boys	Girls
Circles	.10	.25	.06	.16
Patterns	.24	.47	.09	.36
Uses	.32	.38	.20	.28
Improvements	.32	.45	.19	.28
Similarities	.30	.35	.24	.28
Topics	.27	.37	.19	.29
Consequences	.39	.48	.25	.28
Multiple Vocabulary	.46	.53	.17	.21
Rorschach Inkblots	.32	.34	.18	.28
Total D.T. Battery	.51	.63	.29	.42

Table 2 Validity Coefficients of Divergent Thinking Tests from Vernon (1972)

The normal 'first order' correlations are quite substantial for all tests except Circles, and the validities for the total battery of 0.51 for boys and 0.63 for girls are, Vernon suggests, high enough to indicate that "such a battery would be of real diagnostic value to secondary school teachers and counsellors". However much of the prediction is attributable to verbal ability, the residual coefficients dropping to 0.29 for boys and 0.42 for girls. In terms of variance, for boys, verbal ability predicts 22.8 per cent of the criterion and divergent thinking adds 8.1 per cent, and for the girls the corresponding figures are 23.8 per cent and 17.9 per cent respectively.

Commenting on this research elsewhere (Vernon *et al* 1977), Vernon is more critical of the findings than in 1972, noting that when verbal intelligence was held constant "it became more doubtful if the tests are worth the trouble". He also notes that these figures represent confurrent validity and that for longer-term predictions, the superiority of the intelligence tests to available divergent thinking tests might well be greater. The long-term stability enquiry in this study will provide some relevant evidence on this question.

Cronbach (1968) in a re-analysis of the Wallach and Kogan data suggests that when convergent ability is held constant divergent ability adds very little to the prediction of cognitive output. Vernon's (1972) result indicates that this is not altogether true, and advocates of the potential of divergent thinking tests for indicating some 'extra' ability over and above verbal reasoning can be encouraged. It has been pointed out consistently throughout this study that while I.Q. can be regarded as necessary for most creative production, other abilities including divergent thinking could prove to be an important additional ability. The appearance of a significant amount of residual variance in Vernon's study, after the extraction of that due to I.Q., gives some real concurrent validity.

In contrast to the studies of Bennett (1973) and Dewing (1970) with children of 10 to 12 years of age in which little or no sex differences in divergent thinking were found, Vernon (1972) suggests that there are marked differences between the sexes. Bennett's study, however, contained entirely cognitive variables whereas Vernon's study contained a large number of measures of personality and interest. The residual correlations of divergent thinking with the various scores for the 'imaginative story' in Vernon's study were generally significant for both boys and girls, giving support to Bennett's conclusion that divergent thinking is a significant ability in both sexes in performance on that criterion.

Looking more closely at Vernon's non-cognitive variables residual divergent thinking in girls was associated with high teacher ratings for conscientiousness, curiosity, sociability, independence and adjustment, suggesting that teachers are impressed by the divergent thinking girl. For boys the corresponding correlations were generally positive but non-significant. Neither of these results gives support to Getzels and Jackson's finding of a negative relation between 'creativity' and 'likeability', but they are in accord with those of Biggs *et al* (1971). Divergent thinking in girls was strongly associated with artistic talent and literary creativity, and parallel correlations for boys were just significant. The reverse is true for scientific leisure activities.

A "Personality and Attitudes Questionnaire" was constructed by Vernon to measure a number of traits associated with creativity and independence, and yielded eleven different measures including self-concept, need for achievement, independence, and conventionality.

Conventional, rather than unconventional girls, tended to do better on divergent tasks, and positive self-concept was an advantage particularly for boys. Independence was a significant factor in divergent performance, positively for girls and negatively for boys. Rather surprisingly, those boys showing acceptance of school and adults tended to do better than those who were more rebellious.

On the whole, as Vernon concludes, personality patterns were more obscure than those for interests. Both 'interests' and personality variables are to be included in the present study on an exploratory basis and will provide a comparison with Vernon's results.

A number of points of caution however might be sounded. Vernon points out that little stress should be laid on borderline coefficients obtained from about 100 criteria, but concludes that it is reasonable to take note of correlations in related clusters and those that are well above the borderline. When 5 or 6 significant correlations are to be expected at the 0.05 level however it is tempting to invent associations for them.

It is also worth noting that Vernon for the most part follows Wallach and Kogan rather than Torrance in giving multiple options within divergent tests. Thus subjects have to give 'Meanings' for eight patters in 13 minutes, 'Uses' for five objects in 18 minutes, 'Similarities' between six different pairs of objects in 12 minutes, and 'Consequences' for four items in 15 minutes. With this emphasis the divergent thinking score must depend a great deal on verbal ability, even though it is assessed by scoring for originality (number of responses given by less than 5% of the subjects). Verbal and non-verbal scores are added together to form an overall score, though elsewhere, Vernon (1971), in suggesting suitable tests for a battery of tests to operate over a wide age range (from 11 years to college student or adult), maintains that "tests involving non-verbal (drawing) responses, such as Circles and Incomplete Drawings have been conspicuously poor".

The data in Vernon's study was also used to compare performance in formal and permissive testing conditions and half the classes took the divergent tests under conventional conditions and half in a relaxed setting. This comparison is reported elsewhere (Vernon, 1971) and a number of differences between performances under the two conditions were reported. In spite of this Vernon (1972) notes that the "factorial structure of the battery was generally similar under the two conditions" and the data was combined for the 1972 analysis. While this is unlikely to result in strong overall relationships which were not present in both conditions of testing it may result in some blurring of relationships present under more precise conditions.

4.7 SUMMARY

Overall the evidence in this chapter suggests that divergent thinking skills bear some relationship to creative accomplishments though this relationship is not great enough to justify the identification of creative potential simply in terms of divergent thinking. Even if low however, the relationships with creative criteria are generally positive and persist, to some extent, after the effects of general intelligence is controlled. Whether the residual effect of divergent tests is sufficient to warrant their continued development has been questioned, though the variety of scoring procedures, testing batteries, methods of administration and creative criteria makes it difficult to generalise.

The creative criteria often consist of self-report inventories which are consequently not independent of the individual's performance on the divergent tests. Some doubts have been raised about the validity of teacher ratings of creative behaviour, though this is partly aleviated if the teacher has to complete a structured list of creative characteristics rather than simply supply an overall judgement. The utilization of practical criteria has often stopped at an imaginative story, and the use of problem-solving exercises in one study provided evidence of construct, rather than criterion-related validity. Personality characteristics are also likely to affect relationships between divergent thinking and other variables though their mediating effect is not widely reported.

There is obviously a great deal of research still to be done before the value of divergent thinking tests is established. The investigation which follows aims to provide some further evidence, and the experimental procedures are described in the next chapter.

The reliability of divergent tests will then be investigated, scoring procedures, inter-scorer reliability, test-retest reliability, and long-term stability, each being reported on in a separate chapter.

Subsequently the concurrent validity of divergent thinking tests will be looked at in relation to children's creative interests, teacher ratings of their creative behaviour, creative motivation, and an experimental task, the 'Board Game', designed to involve flexible and imaginative aspects of children's thinking. Finally construct validity will be explored with the aid of both cognitive and affective variables including intelligence, academic performance, attitudes, self-concept and personality dimensions.

CHAPTER FIVE
EXPERIMENTAL PROCEDURES

5.1 INTRODUCTION

The practical investigations in this study fall into two parts, one involving a follow-up study, after approximately five years, of 173 pupils previously tested in 1969 and located in the North East of England, and the other based on a population of 176 'local' 11-year-old school children. The main investigation, including scorer reliability, test-retest reliability and validity will be conducted with the latter group, the results from the 15-year-olds being used to investigate long-term stability.

Details of the background and procedures adopted with the 15-year-old sample will be given separately in Chapter 9 when long-term stability is investigated. The same divergent tests, with slight variations, are used in both investigations. The scoring procedures used with the 15-year-old population are identical to those adopted when they were initially tested in 1969 (Richards, 1970), and are similar to those used with the main 11-year-old sample. When giving details of the latter in the next chapter, variations in the 15-year-old marking will be noted.

5.2 THE MAIN 11-YEAR-OLD SAMPLE

The decision to carry out this investigation with children of 11 years of age was made on the basis of a number of factors. The writer had some previous experience of divergent thinking research with children of this age, and a good deal of other relevant research has been based on the studies of Wallach and Kogan (1966) and Getzels and Jackson (1962) with children of a similar age. Children of this age generally possess sufficient verbal skills to deal with the written and spoken framework of the instructions and testing procedures, and yet they have not reached the stage of adolescent development which would make the population less cohesive both psychologically and educationally.

The last year at primary school frequently sees the end of a stage in which educational activities are relatively consistent for all children. Research conducted during this year, however, has an opportunity to involve children, of a wide range of ability, in common activities, and for the researcher to get to know children and teachers in a way which would not be possible in a large secondary school.

In contrast to the view of Wallach and Kogan (1966) that a sample of children from "professional and managerial backgrounds" is likely to be "the one most relevant socio-cultural group for which to establish generalisations", the writer wished to choose a group more representative of the population as a whole. This too is conveniently found in most primary schools in this country.

In order to achieve a reasonably large sample, children were chosen from two schools, one with three 11-year-old classes and the other with two. The larger school was situated on the outskirts of a city in the South West of England, and the other on the outskirts of a small town about 8 miles away.

Both were within walking distance of open countryside and the town centre, and children in either have the opportunity to participate in a wide range of creative activities and interests, both in and out of school. The catchment area of both schools included council and private houses though there was a greater proportion of private houses in that of the larger school. Unlike a previous study of the writer, however, (Richards, 1970) no comparison of the schools is being made and no attempt was made to arrive at a matched pair of schools.

There was no marked contrast in the schools however, and of the five class-teachers four were senior members of staff. The teaching methods in each were predominantly formal. In the larger school two of the teachers were men, and the third was a young woman with about five years teaching experience. In the other, one teacher was male and the other female. It is accepted that there are bound to be wide differences in the children's classroom experience and home background, but the total sample is deliberately heterogeneous in order to arrive at measures of divergent thinking, as well as intelligence, over as representative a sample as possible.

176 children comprising the whole 'fourth year junior' population in the two schools took some part in the testing, complete records finally being available for 161 of them.

Details of the age and I.Q. of the final sample for boys and girls separately, and for the whole sample are given in Table 3.

	I.Q.		Age	
	mean	s.d.	mean	s.d.
Boys (n = 77)	97.39	13.81	11y 4.25m	3.68m
Girls (n = 84)	97.55	14.20	11y 4.54m	3.12m
All (n = 161)	97.47	13.97	11y 4.40m	3.41m

Table 3 Details of the Main 11-year-old Sample

5.3 TESTING MATERIAL

As indicated at the end of the last chapter the testing materials fall into four main groups, the divergent thinking tests themselves, criteria of creative interests and behaviour, convergent tests including I.Q. and school attainment, and a number of scales of attitude and personality. Each of these will now be looked at in more detail.

5.31 Divergent Thinking Tests

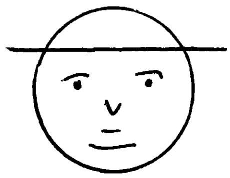
These have already been discussed (Chapter 3) in relation to Guilford's Structure of Intellect Model but are described here in terms of their general use and presentation.

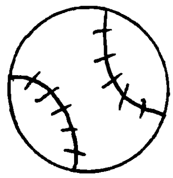
5.311 Circles Test (Torrance 1962, 1974; Guilford and Hoepfner, 1966; Guilford, 1967)

Described by Guilford and Hoepfner (1966), as a test in which figural details are added to several replications of the same basic design to produce a variety of recognisable objects, this test has been extensively developed by Torrance (1962, 1964) for use with school children. It is well suited to group presentation and to children who are slow in their verbal development. It also has a convenient alternate form 'Squares', (Torrance, 1962) which can be presented and scored (with its own frequency distribution of unusual responses) in exactly the same way as 'Circles'. With appropriate substitution of the word 'squares', the following comments on the Circles Test apply equally well to the Squares Test, and separate details of the latter will not be given. Both tests are reproduced in the Appendix.

The subjects were presented with a page containing forty-eight empty circles and asked to make as many different objects as they could. They were told to add lines inside the circle, outside the circle, or both inside and outside, and that they may label the object if they didn't think it would be recognised.

e.g.





tennis ball

The 'test' was labelled 'Circles Game' and attracted a good deal of enthusiasm. From previous experience the time limit of 10 minutes given by Torrance (1974) appeared to be too short and it was felt that a more reliable measurement, particularly of flexibility and originality, would be obtained if the time was extended. Fifteen minutes was therefore allowed and even at the end of this time many of the children wished to continue.

The test is scored for fluency, flexibility and originality and the procedures used are discussed in detail in the next chapter.

5.312 <u>Uses Test</u> (Guilford, 1950, 1951, 1967; Torrance, 1962, 1974)

Since Guilford (1950) suggested this test as a measure of ideational fluency or flexibility, it has been widely used in a variety of forms, from that of Wallach and Kogan (1966) in which subjects were asked to give uses for eight different objects, to the more established form of Torrance (1974) asking for uses for only one object in ten minutes.

With the 15-year-old population the writer has used the form adopted previously (Richards, 1970), with three stimulus objects and a total time limit of fifteen minutes. This was incorporated with the Circles and Consequences tests into a testing booklet and is given in the Appendix. In pilot work prior to the main 11-year-old investigation, however, many children were able to continue thinking of uses for a single object beyond the five minutes allocated, and it was felt that the use of several objects may give insufficient opportunity for less obvious responses to emerge. The format suggested by Torrance, namely one object and a time limit of 10 minutes, has therefore been adopted for the main 11-year-old population.

The children were presented with a sheet of lined paper and asked to write their name and the heading 'Uses for Objects'. As an example, they were asked verbally to suggest uses for a bucket, and the writer wrote on the blackboard five uses, making sure to include, 'to carry water', 'to sit on', and 'to make a helmet'. They were reminded of this when taking the alternate form. They were then asked to write down as many different uses as they could think of for a newspaper, or, in the alternate form, for a spoon. It was made clear that they could write down anything they could think of, no matter how unusual. The responses were scored for fluency and originality and full details of the scoring procedures are given in Chapter 6.

5.313 Consequences Test (Guilford, 1950, 1951, 1967; Torrance, 1962, 1974)

This test, originated by Guilford as a test of fluency of inferences, is put into a less formal context as a 'Just Suppose' activity by Torrance (1962), "to elicit a higher degree of spontaneity and to be more effective with children". A similar format is adopted here, though the time limit of five minutes allowed by Torrance is increased to ten minutes.

The test was introduced by an example. "I want you to imagine that the following thing suddenly happened, and to think of what other things would happen because of it. Just suppose that we only had one arm. What things would happen as a result?". The writer then listed five responses on the blackboard including the following:

(i) we couldn't use a bow and arrow
(ii) couldn't thread a needle
(iii) only need one glove.

The children were then asked to write down as many different things that might happen 'if we had no hair on our heads', and for the alternate form 'if we did not need to eat'.

With the 15-year-old group both items were included in the same time limit of 10 minutes, but once again, in order to give time for more 'searching about to be done', only one item was included in the test for the main 11-year-old sample. In the interests of simplicity, the second item was also reduced from the more comprehensive situation of 'no need to eat or drink' to just 'no need to eat'.

The responses were scored for fluency and originality as described in Chapter 6.

5.314 <u>Representative Nature of the Tests</u>

The divergent thinking tests therefore comprise two verbal and one figural test and provide a group of test which should enable a comparison to be made with other researches, and previous work of the writer (Richards, 1970). All three tests have figured prominently in previous researches, stemming from Guilford's early development of materials for use in his factor analytical studies of creative thinking Guilford *et al*, 1951), and from their development in a form more suitable for children, in Torrance's batteries of creative thinking tests both in his outline version of the Minnesota Tests of Creative Thinking (Torrance, 1962) and in the latest issue of the 'Torrance Tests of Creative Thinking' (Torrance, 1974). In addition to their wide use within Torrance's testing battery numerous other studies have used their own version of one or more of these tests. Vernon (1971), for example, includes all three amongst the seven tests in his study of the effects of administration and scoring on tests of divergent thinking. Hudson (1966) refers to the Uses test as almost the conventional divergent thinking test, and considers it "perhaps the most convenient and versatile of open-ended tests". He employs it as one of the two tests he uses to distinguish convergers and divergers. Dewing (1970) selects the Uses test and the Consequences test as the two tests on which to base her study of the reliability and validity of selected tests of creative thinking, and Debney (1969) in a study of the relationships between 13 creativity tests found the same tests to be among those most clearly related to an internal validity criterion - high and low groups defined by the battery of tests as a whole.

The Uses test is one of the tests employed in the major investigations by both Wallach and Kogan (1966) and Getzels and Jackson (1962); and Hasan and Butcher (1966) add the Circles test to the divergent battery in their partial replication of Getzels and Jackson's study. Versons of the Uses test and the Circles test are also included in studies by Haddon and Lytton (1968, 1971) and Lytton and Cotton (1969) into the effects of different teaching approaches on the development of divergent thinking abilities, though in these studies the Circles test showed some different patterns of results from those of the other tests. Child (1968) on the other hand found that in factor analysing a large battery of tests, including seven of divergent thinking, a divergent factor of ideational fluency was formed by the three tests chosen here, namely Uses, Consequences and Circles.

Inconsistencies in the results of studies using the Circles test are noted by Vernon (1971) and consequently, though he recommends the inclusion of both Uses and Consequences tests in forming a divergent battery, he does not advice the inclusion of test with non-verbal responses such as Circles. Evidence presented by the writer (Richards, 1970) would tend to support Vernon's assessment of the performance of the Circles test, as it showed some definite factorial separation of the Circles test from other tests including verbal tests and figural tests with verbal responses. The exclusion of tests with non-verbal responses however, while it might make the divergent tests more homogeneous and be advisable if a summed score was required, would limit the scope of any investigation into tests of divergent thinking.

The three tests having been selected it was encouraging to find all three recommended by Hargreaves and Bolton (1972) in their investigation into the problem of selecting creativity tests for use in research.

5.32 Criteria of Creative Interests and Behaviour

5.321 Teacher Ratings of Creative Behaviour: 'Indicators of Creative Behaviour Rating Scale'

The testing took place in June/July towards the end of the children's school year, and they were well known by their class teacher who had taken them for that year. It was felt that the class teacher would therefore be able to give valuable judgements about the extent to which the children had exhibited creative characteristics during that year.

In suggesting non-test ways of identifying creative behaviour Torrance (1967) reports on a number of descriptions of creative children made by parents, teachers, counsellors and administrators. In one investigation eighty-seven members of a creative thinking seminar group each contributed a list of five behaviours which they felt were the best indicators of creative talent. The most frequent types of behaviour suggested were compiled by Torrance into thirteen categories, from "curiosity, inquisitiveness, investigativeness, penetrating questioning, etc.", given by 66% of the participants; to "daydreamer, pre-occupied, etc.", given by 10%.

From the descriptions given, the writer has drawn up a list of ten characteristics to use as a basis for the teacher's ratings in this study. Teachers were presented with this list together with a recording sheet giving the names of the children in their class, and were asked to rate each child on each of the characteristics, entering 2, 1 or 0 if the child had, during the year, shown the characteristic 'to a great extent', 'sometimes', or 'very weldom or not at all', respectively. The total score for each child, with a possible range from 0 to 20, is used as the index of creative behaviour.

The rating scale as presented to the teachers is reproduced in the Appendix.

5.322 Creative Leisure Activities Checklist: 'Things you have done in your spare time'

This is a checklist of 95 creative activities practiced by children in their spare time. They are specifically asked to include only the things they have done for themselves, not things they have been given or told to do. Originally compiled by Torrance (1962) and reported on in a validity study by Dewing (1970), it "includes activities related to language arts, science, social studies, art, and other fields" (Torrance, 1962). Of the original 100 items the writer has retained 95, with slight change of vocabulary, to suit British school children. A copy of the checklist is included in the Appendix. One mark is awarded for each item ticked. In attempting to assess children's interests and activities there is always a problem, as Evans (1965) points out, of differentiating between the level at which an interest may function and the simple choice of interest. The present inventory is of the latter type and will be interpreted as such when the results are presented. Limitations in the use of this type of assessment will also be considered.

5.323 Interests

Following discussions about their interests with groups of 9 to 11-year-old British school children, Barker Lunn (1970) constructed a 30-item 'Interests' questionnaire specifically including items judged to have "creative content". The scale was subjected to a considerable amount of study including factor analysis and two consistent sub-scales "both concerned with creative interests, one imaginative and the other logical" were established. The former consists of seven of the items and the latter, four.

The remaining nineteen items give general information about the activities children "like doing either in or out of school". As such they might appear to give a further index of leisure activities that might be related to divergent thinking ability. A large number of these items however include games and 'passive activities' such as playing football, cricket, playing marbles, climbing trees, watching T.V. and going to the cinema, and are not relevant to the thinking abilities being investigated. Seven of the items however, drawing, collecting stamps, writing a daily diary, doing crossword puzzles, sewing, gardening and dancing are similar to items on Torrance's checklist, and were felt worth including in an overall 'creative interests' measure, which will also provide some comparison with the Creative Leisure Activities Checklist described above.

The children are asked to indicate their feelings about the interests by marking each activity in the following way:

Like very much = 2

Quite like = 1

Never tried it or don't like it = 0.

Their score for each of the two sub-scales, and the total score for all the interests except those designated as sporting/passive were recorded, and will be used as criteria for the investigation of concurrent validity.

A copy of the scale is given in the Appendix. Items forming the total and sub-scale scores are lettered accordingly, the 'imaginative' interests are marked 'I', the 'logical/ analytical' interests are marked 'L' and the 'general' active interests are marked 'G'.

5.324 Creative Motivation Scale: 'Things I would like to do most'

This scale, amended from Golann (1962) and Dewing (1970), requires the children to choose one preferred activity from pairs of creative and non-creative items judged to be equally socially desirable.

The amendments are chiefly those of language, found after a pilot study to be inappropriate. Thus 'Dress the windows in a bookshop' (Q.6) is changed to 'Arrange a display in a bookshop window'; 'penmanship' (Q.8) becomes 'handwriting', 'store' (Q.18) becomes 'shop'; and 'the great books' (Q.20) becomes 'famous books'. The choice in question 17 is changed from 'Do a neat careful report' or 'Do an original report', to 'Write an account of a visit' or 'Write a story'.

The scale is reproduced with scoring key in the Appendix. The more 'creative' activity is denoted by the appropriate letter on the right hand side of the questionnaire. One mark is awarded for each creative activity chosen.

5.325 'The Board Game'

5.3251 Introduction

This is an experimental task designed by the writer to involve productive and imaginative thinking for a 'good' solution. It incorporates several 'Gestalt' principles and it is hypothesised that children's divergent thinking abilities will be a positive factor in effecting a solution.

As a result of previous research into children's mathematical ability and the effects of different teaching approaches, the writer had felt that written test materials were unlikely to do full justice to the type of thinking ability fostered by some of the newer approaches to mathematics teaching. Although there was some slight evidence that such methods might have positive effects on the productive and original aspects of children's thinking, it was suggested that "further studies in practical open-ended situations or exploring the type of problem described by Wertheimer (1961) are necessary to substantiate this hypothesis" (Richards and Bolton, 1971). A similar consideration applies to evaluating the validity of divergent thinking abilities. Finding a suitable practical exercise however was not easy. Even the problems designed to exercise thinking skills, by researches such as Wertheimer (1949, 1961) and Bartlett (1958), are predominantly in a symbolic format and involve a considerable amount of logical reasoning. The 'lateral thinking' exercises of De Bono (1967, 1971), and certain other problems with a practical flavour such as matchstick problems, pentominoe puzzles, or arrangements of counters, were tried out in a pilot study but they did not sustain the motivation of the children.

Too often there is an unique solution to such problems and attempts have to end with 'all or nothing'. If they don't 'see' the solution quickly, many children lose interest, especially if they view the problem as some sort of 'mathematical' puzzle which they 'know' they can't do! Faced, for example, with making four equilateral triangles out of six matches they frequently declared the problem 'impossible'.

On the other hand many of the problems had been met with previously and the solutions - such as building six matches into a regular tetrahedron, were often already known. This made it difficult to judge whether the solution was a personal achievement or not.

As a result it was felt necessary to get away from a 'puzzle' framework into a practical exercise with more scope for experiment. As discussed in Chapter 3 the early Gestalt approach to the study of problem solving (Maier, 1930, 1931; Duncker, 1945) utilized a number of more practical problems and these, together with the type of productive thinking described by Wertheimer (1949, 1961) led the writer to design a practical experiment to be referred to as the 'Board Game'.

There is no unique solution, but for a successful outcome the solver needs, as in Duncker's (1945) problems, to be willing to experiment, to look at things in new ways, and to be able to change direction in his thinking. He has to arrive at a "combination and organisation of elements" which meets the required condition (Maier, 1931); and to achieve this he has to think of unusual uses for objects rather than see their function as 'fixed' by their habitual use.

5.3252 <u>Description and Administration</u>

As a prerequisite the subject is assumed to understand the nature of an electric circuit and to know that all metals conduct electricity. This formed the subject of classwork done prior to the Board Game, and will be described under 'Organisation' later in this chapter. Having had this experience the subjects were individually presented with the Board Game, set up exactly as shown in the photograph (Figure 4).

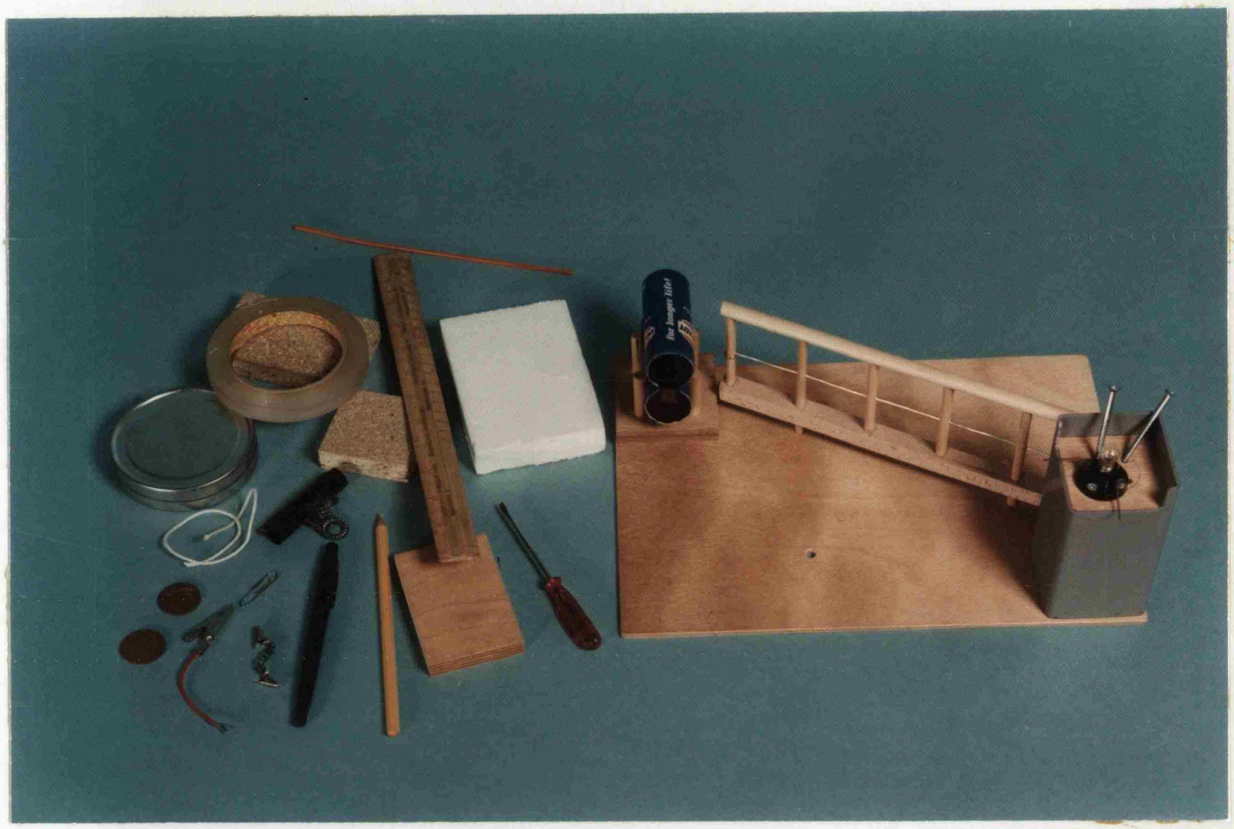


Fig.4. Initial Position of the Board Game

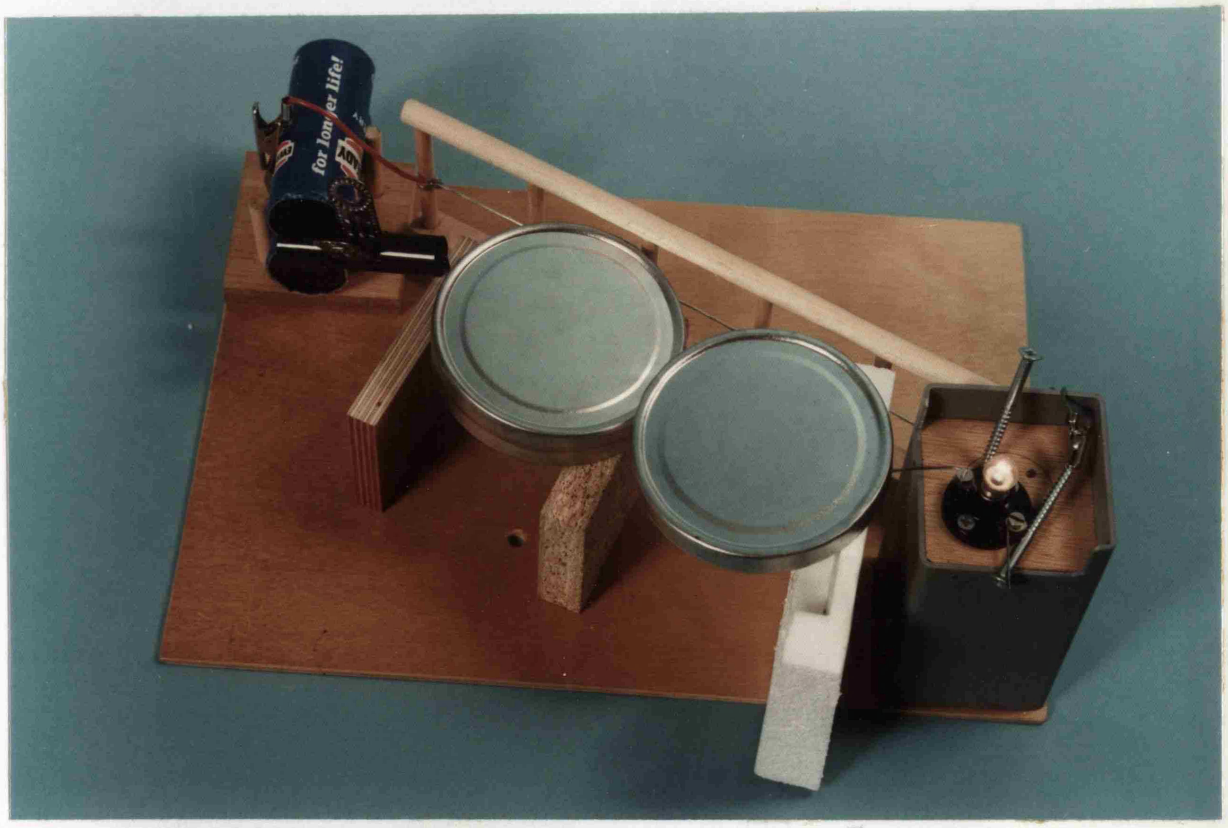


Fig.5. An α-solution to the Board Game

A good 'α' solution is also shown in the adjacent photograph (Figure 5), and will be useful for reference during the following description. Photographs of all the main types of response to the problem are included in the Appendix.

The board itself, which measures 33cm x 23cm, carries a plastic tower surmounted by a bulb in a bulb-holder, a fence and a battery holder. The problem is to light the bulb at the top of the tower with the battery placed at the other end of the board. Anything on the table can be used to complete the circuit.

Each child was introduced to the game individually and informally, though the same information was given and the same procedure carried out each time. The introduction proceeded as follows:

"Do you remember the lesson we had on electricity?"

"What did we find out about all metal objects?"

("They all conduct electricity")

"Here is a bulb on top of this tower, and here is a battery."

(Handing over the battery.)

"Show me how to light up the bulb; you can use this piece of wire to complete the circuit."

(An extra piece of wire was needed to join one terminal of the battery to one at the bulb. The other terminals can be connected directly with wire protruding from the bulb-holder.)

"Good"

"Now I'm going to put the battery down at this end in the holder"

"I want you to make a circuit to light the bulb again, with the battery at this end of the board."

"You can use anything you like on the desk/table."
"Just a quick flash on the bulb and you've done it."
"Would you like to try?"
"You can use anything you like."

The materials available to the children can be seen in Figure 4. Some of the objects would conduct electricity, others would not. Non-conducting materials could be used for support, however, or in any other way that was thought to be useful. The objects which conducted electricity were as follows:

paper clip	bulldog clip
coil of wire (crocodile clip at each end)	two 2p pieces
	clip on the pen top
coil of wire (crocodile clip at one end only)	wire on the fence
round tin with lid	wire attached to bulb holder
two screws	screwdriver

Other objects provided were:

coloured pencil	roll of sellotape
ruler	drinking straw
3 blocks of polystyrene	piece of string
3 blocks of wood	

5.3253 Development and Assessment of Performance

A number of pilot studies were conducted with the apparatus, the above description being that of the final product. It was found that most children began by uncoiling the wires, found that they were far too short to be sufficient on their own, and gradually introduced other objects into the circuit. Some of the most intelligent children failed to complete the circuit, while others solved the problem in less than 10 minutes. Allowing up to about 35 minutes per pupil, about a half of those studied succeeded in lighting up the bulb in one way or another.

In order to arrive at an 'elegant' solution it is necessary to modify some objects, and to incorporate into the circuit others that might not at first sight appear to be useful. Unbending the paper clip helps, so does using the screws which appear loosely built into the board, and a more advantageous position can be found for the battery whilst still keeping it in the holder. While some children carried out these operations without apparent inhibition, others had to be reassured, and others said later that either they "just didn't think", or "didn't think they could do that". Whenever asked for permission to modify the board or the objects the experimenter consistently adopted the response of "You can do anything you like".

The coins, penclip, screwdriver and bulldog clip can all be useful in building the circuit through the tin, especially if opened, and the wire strand on the fence are of major importance. Solutions can be obtained without using all of the objects, in fact the best solutions often use the least number of bits and pieces.

The difference in height between the bulb and the board means that blocks of wood or polystyrene can be useful in building up the circuit to the right height. Some children managed to reach the tower but could not bridge the final gap from board-level up to the bulb. Others lifted the fence, with its horizontal wire, out of the sockets in the board and made it slope gradually up to the bulb.

A study of the responses of each child during the pilot study encouraged the writer to believe that the most successful solutions needed more than logical reasoning, and that the Board Game was a feasible practical exercise to involve aspects of creative thinking as discussed in Chapter 3. It is suggested, therefore, that the most successful solver will be the person who is able to generate a large number of ideas, is flexible enough to put them into operation, and at the same time is able to view the different elements in relation to the problem as a whole. These qualities are also reflected in divergent thinking performance and it will be hypothesised later that divergent thinking will be a significant factor in children's performance on this task.

Children's responses are similar in a number of ways to those observed by Wertheimer (1949, 1961). In presenting his productive thinking tasks he found 'fine, sensible and original processes' which he termed α-responses, and also 'blind, slavish and unsuccessful' attempts which he termed β-responses. In the same way here, some children approached the task in a piecemeal fashion, joining up anything that came to hand, laboriously following the obvious method that they first thought of. They often continued in this way for up to thirty-five minutes before they gave up or the experimenter intervened to help them to a solution. All of the children, in fact, were helped to achieve some sort of success. This was done, however, not by using the experimental material, but by adding two lengths of wire to the circuit or by placing the battery nearer the bulb.

The successful solvers usually adopted more flexible approaches and appeared to have a deeper structural view of the situation involving changes in the functional meaning and grouping of the items. While some children therefore managed to achieve a good 'gestalt', others remained blind to context, unable to appreciate the overall structure of the problem.

The β-responses, Wertheimer suggests are typical of the child who rejects the problem with the comment that "We haven't done this before" or "We haven't learned how to do this". All the children in the present investigation, however, had achieved some success in the electricity lesson, and all made some token attempt. Some of the children, nevertheless, were reluctant to commit themselves practically, prefering to 'think about the problem' and rarely went beyond the stage of utilizing the obvious pieces of wire with their attached crocodile clips. Their reluctance to take risks is a characteristic which may also be reflected in their divergent thinking performance.

A scheme for assessing the quality of the children's performance was built up on the basis of Wertheimer's α and β-responses but with an extension to six different categories. They do not define completely distinct types of solution, but they reflect the different types of performance that could reasonably be distinguished.

5.3254 Categories of Performance

The following descriptions of children's responses are used to classify their performance:

α-responses

Solves quickly, modifying objects if necessary, and building up with a fine overview of the problem. Sees the relationship between overall structure and individual properties of the objects. Time: less than 10 minutes.

β_1-responses

Solves the problem after quite a lot of experiment, but prepared to change direction of approach, and final solution uses the objects efficiently with modifications (similar to an α-solution). Time: approximately 10 to 15 minutes.

β_2-responses

Solves the problem, but in a fairly laborious, straightforward way, using just about all the articles in a linear fashion. Some modifications. Time: approximately 15 to 25 minutes.

β/j-responses

Final solution is unstable and laborious but eventually gives a successful outcome possibly with help in holding together. If the circuit is technically complete but fails to conduct properly because of its instability, slight help can be given to hold together. This help can be given at any time after 20 minutes if the pupil indicates that the circuit is complete but won't light the bulb, and that he is glad to accept help rather than try to modify the circuit further. Time: approximately 20 to 35 minutes.

j_1-responses

Works for a long period of time, tries quite a lot of objects, links them up inefficiently, can't jump the final gap. Gives up. Doesn't see the whole problem, but views it only in terms of joining things together and hoping they will reach. Time: approximately 20 to 35 minutes.

j_2-responses

Spends time unconstructively, possibly thinking but without much construction. Seemingly unable to grasp the situation of the problem as a whole though understanding the elements of an electric circuit. "Can't see how to do it." Gives up sooner or later without much practical result. Time: approximately 10 to 25 minutes.

A photograph of a final solution or attempt in each of these categories is given in the Appendix.

In order to arrive at these categories details of each child's responses were recorded on a prepared recording sheet. Notes were made on articles used and whether or not they were modified, and on the method of approach. Certain actions such as changing the position of the battery, or changing round articles to improve the solution, called for specific comment, and other comments were made, with a note of the time, at regular intervals. A drawing of the final solution or attempted solution was also made. A copy of the record-sheet 'Responses to the Board Game' is given in the Appendix.

In some of the subsequent analysis a numerical score is adopted for each of these categories, from 5 for an α-solution to 0 for a j_2-solution. Further details of the arrangements for administering the Board Game in the present study, and of the preliminary lesson on electricity, are given later in this chapter under 'Organisation'.

5.33 <u>Convergent Thinking</u>

5.331 <u>Intelligence</u> (V.R.Q.)

Two parallel forms of the Moray House Verbal Reasoning Tests had been given to the children in both schools in completing the County's personal record procedure for Junior Schools, and the results were made available to the writer for confidential research purposes. The tests are standardised to a mean of 100 and a standard deviation of 15.

In the interests of the factor analysis to follow, a balance needs to be kept between different groups of tests and both I.Q. scores are therefore included in that analysis. For all other purposes the I.Q. is taken as the result of the first testing. Although a more reliable measure would be obtained by averaging the two scores, this would in turn affect the reliability and validity of the divergent thinking tests when the effects of intelligence are controlled.

5.332 Academic Performance in English and Mathematics

Overall ratings of the children's attainment in school work were obtained from class teachers for English and Mathematics. Each teacher had already compiled school end-of-year grades for these subjects, again for record purposes, but were also asked to allocate a rating for each child on a common scale presented by the writer: unlike the creativity ratings, these were not therefore basically subjective, but were based on the children's work in classwork and tests throughout the year.

The grades suggested were:

A	B	C	D	E
Very good	good	average	weak	very weak

A prepared class list with columns for the appropriate grade was given to each teacher, accompanied by an introductory sheet, 'Children's School Work and Personality' is given in the Appendix. It also outlines the request for ratings of certain aspects of the children's personality.

5.34 Personality and Attitudes

5.341 Junior Eysenck Personality Inventory

Details of this scale, designed to measure the two major personality variables of neuroticism or emotionality, and extraversion/introversion in children are given in the test Manual (Eysenck S.B.G., 1973). It is designed for children from 7 to 16 years of age, though some of the least able of the 11-year-old children in the present study required assistance in reading some of the items.

In addition to the Neuroticism scale 'N' (24 items) and the Extraversion scale 'E' (24 items), the test also provides a Lie scale 'L' (12 items). The latter measures the extent to which subjects 'fake' responses for one reason or another. Although the term 'Lie scale' has a somewhat unfavourable flavour, it is often associated with the subject's attempt to present himself in a good light (Michaelis and Eysenck H.J., 1971), and can provide a measure of test-taking attitudes or "conformity to social pressures" (Eysenck S., Nias and Eysenck H.J., 1971). It could provide interesting information in the exploratory section of this study and all three measures, E, N and L are retained in the analysis.

5.342 Self Concept: 'What I am Like'

A 'Self-Appraisal Scale' designed by Davidson and Greenberg (1967) for use with children aged 10 to 11, was chosen to assess the children's self-concept. The introduction to the scale was slightly amended by the writer from the American presentation, and the scale was entitled 'What I am like'.

The test consists of 24 items each giving a concise description of what children are like from 'neat', 'a big help at home', 'nice looking' and 'polite', to 'lazy', 'careless', 'nervous' and 'sad'. Subjects are asked to tick opposite each item to show whether they think they are that way, 'most of the time', 'about half the time' or 'hardly ever'. These categories are scored 3, 2, 1 respectively for favourable items, and in the reverse order for unfavourable ones. The total score is a measure of the children's self-concept.

A copy of the scale is given in the Appendix.

No problems appeared in administering or marking, though the 'I THINK I AM:-' format led one slightly dreamy boy to respond 'Robert Smith'!

5.343 Teacher Ratings of Children's Personality and Pleasurability

When responding to the request for grades on children's attainment in English and Mathematics, teachers were also asked to rate the children in their class on two questions of personality, 'shy?' and 'lively?', and also to what extent they were 'a pleasure to have in class'. They were given a recording sheet with the children's names already inserted, and asked to use the following grades

A	B	C	D	E
Very often	quite often	sometimes	seldom	not at all

A copy of the introductory sheet 'Children's School Work and Personality' is given in the Appendix.

5.344 Attitudes to School and School Work

Noting that no suitable measures of pupil's attitudes were available for use in her large-scale research study into the effects of streaming and non-streaming in Junior Schools, Barker Lunn (1969) constructed a set of 'Primary Children's Attitude Scales' for use with children aged 9 to 11 years. These provide a useful set of resources for other research workers, a considerable amount of evidence for their reliability and validity having been accumulated in an analysis with over 2000 pupils.

A specimen set and 'Manual of Instructions' was supplied by the N.F.E.R. for use by the writer. Of the ten possible scales incorporated in the 64-item questionnaire, six were scored for this investigation. The following summary will enable these scales to be scored, and a copy of the questionnaire is included in the Appendix, but anyone envisaging using the materials should apply to the N.F.E.R. for the complete manual.

The original cover sheet, with the N.F.E.R. heading, was amended into a less formal cover sheet for present use, under the title 'What I think about School'. All the instructions and guidelines were preserved however, including the note regarding confidentiality between researcher and pupils, and in presenting the material, these were read aloud to the class while they followed the wording in their own booklet.

The following scales are included in this study:

(i) Attitude to School and Intent in School Work

It is possible to score these two aspects separately though they intercorrelate highly and the writer has adopted Barker Lunn's (1969) suggestion that there is a case for combining them. The scale therefore includes both general attitudes, such as 'School is fun' and 'I would leave school tomorrow if I could', and statements directed towards school work such as 'I enjoy most school work' and 'we spend too much time doing arithmetric'.

There are 12 items altogether 18, 25, 41, 48, 56, 60, 8, 12, 23, 31, 47, 64, the first six emphasising general attitudes to school and the others school work.

The scores to be allocated to each item are noted on the specimen questionnaire in the Appendix. Half marks are awarded for 'contrived items' which demand two items being endorsed positively before credit is given. Any odd half marks are consequently rounded *down*. Total scores can range from 0 to 12, and high scores, as on all scales, represent positive attitudes.

(ii) Importance of doing well

Items on this scale stress the child's attitude towards achievement, such as 'I work and try very hard in school' and 'Doing well at school is most important to me'. There are five items in all, 11, 43, 44, 52 and 63, and each is scored 2, 1 or 0.

(iii) Conforming versus non-conforming

This scale covers these two opposing types of behaviour, with items such as 'I dislike children who are noisy in class' and 'It's nice to fool about in class'. There are five items 3, 6, 22, 33 and 35, and scores range from 0 to 5. Half marks are rounded down.

(iv) Relationship with Teacher

These items emphasise the extent to which the child thinks the teacher shows concern for him, rather than his liking for the teacher. Items include 'Teacher is interested in me' and 'Teacher thinks I'm a trouble-maker'. There are six items 4, 24, 36, 49, 51, 54, scored in the way indicated. Negative half marks are added algebraically and odd halves are again rounded down. If there are two marks of '$-\frac{1}{2}$' however the total algebraic score is *increased* by one, (or alternatively the negative marks are ignored). This is necessary to prevent the scoring for a pair of contrived items affecting the swing of other single items.

(v) Anxiety in the Classroom Situation

Items on this scale refer to fears and worries in the classroom, for example, 'I would feel afraid if I got my work wrong' and 'Children who can't do their school work feel ashamed'. The relevant items are 20, 27, 28, 32, 38, 58, 61 and possible scores range from 0 to 6. The scoring is slightly different from the other scales, some items being credited with ticks instead of marks. One mark is awarded for each of the following: if *both* items 27 and 28 are ticked at least once, if *both* items 27 and 38 are ticked twice, if *both* items 38 and 61 are ticked at least once.

(vi) Academic Self-image

This scale reflects children's image of themselves, not in general terms as in the Self-Concept (S.A.S.) scale, but in terms of school work. Thus items include 'I'm useless at school work', 'I'm very good at sums', and 'I find a lot of school work is difficult to understand'. There are nine items, 5, 7, 14, 17, 30, 42, 45, 50, 57, and the total score can range from 0 to 18.

5.4 ORGANISATION

5.41 General Organisation

Although it might have been possible to have carried out the pencil and paper tests during a number of visits to each school, the time needed to look at each pupil's individual performance on the Board Game made a more permanent attachment necessary. The writer also wished to keep the testing sessions as informal as possible and consequently needed to avoid cramming too many tests into one session. These considerations are particularly important in attempting to provide the right atmosphere and gain children's cooperation for tests of divergent thinking, intents and attitudes.

The actual testing time was therefore spread over five weeks, during which time the writer and an assistant were continually involved in either group testing or carrying out the Board Game experiment. The assistant was a recently qualified student of the writer's who had just gained his Postgraduate Certificate in Education. He had been involved in some of the pilot work developing the Board Game, and after a short introductory period was able to supervise its administration and assessment in cooperation with the writer. The arrangements for this will be returned to later.

The testing materials, apart from the Board Game were given in a group format to each class in their normal classroom. Arrangements had to be made with each teacher, on a day to day basis, to avoid interruptions due to such activities as swimming, end-of-year plays and musical rehearsals. The order in which the tests were given however, and the length of each session were maintained across all the classes. It was not possible within the school timetables to keep the classes exactly in phase, sometimes a class would have a day without any tests, and occasionally would have one session in the morning and another in the afternoon. At no time was more than one session given in a morning or an afternoon. The lack of phase had the advantage of making it difficult for children to realise that a definite pattern existed across the classes, and would have reduced the effects of communication between classes. The latter no doubt took place, but apart from performance on the Board Game, passing references were considered unlikely to alter children's general level and pattern of responses.

To help avoid any meaningful communication about the Board Game, it was kept in a separate part of the school and packed away between sessions. While children may have described its general nature and appearance to others their exact performance would have been difficult to describe. To complicate this further the children, on finishing the Game, were congratulated on their effort and asked not to tell anyone else how to do it, "as I want to see how they do". The congratulations, and the suggestion that they had found a clever method, were given to all the children even for the most unlikely solutions. In this way it was felt that any subsequent hints to other children, who had not even seen the Game, would not be particularly helpful.

5.42 Testing Sessions

The tests were given in eight different sessions, the retesting with alternate forms of the divergent tests taking place one week (± 1 day) after the original testing. Apart from the divergent thinking tests, there were no time limits, though a rough indication of the time taken is given in the following list of the eight sessions:

1. 'What I Think About School (20 mins.)
   Interests (5 mins.)
2. Circles Test (15 mins.)
3. 'Things You Have Done in Your Spare Time (15 mins.)
   Uses Test (10 mins.)
4. 'Things I Would Like to do Most' (10 mins.)
   Consequences Test (10 mins.)
5. Junior Eysenck Personality Inventory (15 mins.)
   Self-Concept (5 mins.)
6. Squares (15 mins.)
7. Uses Test (10 mins.)
8. Consequences Test (10 mins.)

5.43 Preparatory Lesson and the Board Game

Before embarking on the testing sessions the writer spent half a morning with each class, giving a lesson on a simple electric circuit and things that conduct electricity.

The lesson took the form of a class experiment each pair of children being given a 'kit', as shown in Figure 6, made up as follows:

Cycle lamp battery

Bulb

Two wires with crocodile clips

Plastic tube	Perspex	Nail
Metal clamp	Copper Wire	Hacksaw blade
Piece of rubber	Screw	Cardboard
Foam ball	Polystyrene	Metal plate
Stone	Button	Formica

Wooden rod	Straw	Plastic covered wire
String	Plywood	Wool.

Fig.6. Basic Kit for the Electricity Lesson

Using the battery, bulb and wires, the children were shown how to complete a simple circuit, and how to incorporate an object into the circuit. They tested the objects given and any others they had at hand and listed them as 'Conductors' and 'Non-conductors'. Their findings were discussed and finally written down by each pupil. The main conclusion, that all metals conduct electricity was stressed, though also that some objects, such as the metal plate, will not conduct if covered with paint or dirt. It was also noted that electricity can be conducted through water and other substances, such as the carbon lead in ordinary (not coloured) pencils. A typical 'report' on the experiment by one of the children is reproduced in the Appendix.

Having completed this lesson the children were in a position to attempt the Board Game, and each school allowed the children to be drawn out of their normal lessons to participate in the experiment. The schools also provided some space for the experiment to be conducted in isolation from the other activities. In one school this was a handicraft room not being used a great deal at that stage of the year, and in the other an unused cloakroom.

Four sets of the Board Game were set up well away from each other on double desks or tables, and the writer, with the help of his assistant, found it possible to supervise four children at a time, each child individually working on his own apparatus. A record sheet was completed for each pupil, two being filled in by the writer and two by his assistant. It was possible for the experimenters to move freely about the room however, and to communicate to each other if any special features needed discussion. The final solutions were left standing until all four pupils had finished and were then discussed by the experimenters in conjunction with the record sheets. Each solution was finally placed in one of the performance categories outlined earlier in this chapter.

Each of the testing sessions on the Board Game was conducted in this way, the apparatus being set up in a standard format as shown in Figure 4, and any damaged articles replaced. On average, each session took about 40 minutes, and testing took place whenever time was available between the group testing sessions.

CHAPTER SIX
THE SCORING OF DIVERGENT THINKING TESTS

6.1 INTRODUCTION

It has to be appreciated that in comparison with the automatic scoring procedures that can be applied to standardised tests whose questions demand specific correct answers, the scoring of tests of divergent thinking will usually involve tedious and often complex procedures. The more encouragement the tests give for a subject to think divergently and depart from the usual, the more difficult it is to score his responses reliably whilst at the same time crediting what Getzels and Jackson (1962) term the "richness and uniqueness" of a subject's response.

The researcher is thus faced with the problem of devising a marking procedure that will maintain a balance between objective and subjective assessments and at the same time preserve reliability and validity. What is more the procedures need to be realistic in terms of the time and effort they involve. As Vernon (1971) observes the problem of scoring is a preliminary difficulty that must be overcome in using divergent thinking tests though, as he says, this is seldom referred to in the literature.

Most attempts at devising scoring procedures for divergent thinking tests have been based on the creative thinking abilities of fluency, flexibility and originality hypothesised by Guilford (1950), and subsequently formulated into his Structure-of-Intellect model (Guilford, 1956). Although, as noted in Chapter 3, Guilford's model of the intellect is not without its critics (e.g. Eysenck, 1967; Vernon, 1964), his conceptualization of the divergent production abilities within his Structure-of-Intellect, together with tasks designed to assess them, has provided a springboard for other attempts at devising materials for measuring divergent thinking. The most thorough attempt to establish a battery of such tests is found in the work of Torrance (e.g. 1962, 1965, 1972) culminating in the latest versions of the Torrance Tests of Creative Thinking, with accompanying Norms-Technical manual, Directions manuals and Scoring guides (Torrance, 1974). As noted earlier Torrance's materials have been used in a very large number of studies and they provide the most comprehensive set of procedures available for the administration and scoring of the tests used in the present study. Although a number of criticisms of his materials will be made, the value of Torrance's work in developing instruments and procedures for use in investigating divergent thinking abilities must be acknowledged.

Goldman (1967) points out that the divergent thinking tests used in Guilford's early studies were designed to assess only one specific factor in his Structure-of-Intellect model, and were consequently scored for that factor only, whereas Torrance, in his development of the tests has incorporated several factors into one test. Although a number of researchers (e.g. Hudson, 1966; Wallach, 1970; Hargreaves and Bolton, 1972) consider that the time and effort required to calculate anything more than fluency scores do not justify the small amount of extra information gained, a single test scored according to Torrance may yield scores for fluency, flexibility, originality and elaboration. Few independent experimenters however, are likely to score a single test for all these abilities, and it would be inappropriate to do so, as indicated below, under certain instructions for administering the tests.

Although he gives a guide to scoring both his verbal and figural tests for elaboration, Torrance (1974) notes that some difficulties have been experienced in obtaining inter-scorer reliability of a high order with the verbal tests, and cautions potential users that "Until more data are obtained concerning the Verbal Elaboration Score, most users may prefer not to use this score".

In administering the figural tests the elaboration ability is encouraged by an instruction to the pupils to keep adding to any idea they think of. In particular, in the Circles Test, Torrance encourages pupils to make as many different objects as they can and to put as many ideas into each one as they can. The latter part of this instruction was not included in administering the tests in the present study and no elaboration score was extracted. Scoring methods will however be scrutinised in relation to the three abilities of fluency, flexibility and originality.

In spite of the work of Torrance, it is typical of the lack of uniformity in scoring methods for divergent thinking tests that even these three abilities are frequently scored in different ways. It is not difficult to establish from most researches the nominal abilities for which the tests are scored but frequently scant attention is given to the precise scoring procedures used. This makes replication of experiments difficult and adds to the confusion that, as Yamamoto (1965) observes, has been characteristic of experimental approaches to creativity.

6.2 THE REPORTING OF SCORING PROCEDURES IN DIVERGENT THINKING RESEARCH

The following illustrations reveal some of the lack of detail and are typical of the way the scoring procedures are treated in many reports of research involving tests of divergent thinking.

Ogilvie (1974), in an investigation of creativity and curriculum structure, notes that his tests owe much to Guilford's and Wallach and Kogan's conceptions with regard to both type and task administration, and that they include measures of complexity, associative fluency, and originality. When information about the tests is given however no further information on scoring is provided beyond the comment that "complexity and originality scores can be constructed".

In Hasan and Butcher's (1966) replication with Scottish children of Getzels and Jackson's study, it can be assumed that the scoring procedures used with the tests, which were common to both experiments, were as outlined by Getzels and Jackson. A number of other tests were used however, with no scoring details except a note that the creativity score was the aggregate of "scores obtained from ten tests of creativity". The inclusion of the Circles Test alone could have resulted in a further four scores, for fluency, flexibility, originality and elaboration, or it might have simply been scored for fluency. In the reference that Hasan and Butcher give for the Circles Test, (Torrance, 1962), Torrance himself describes two versions of the test, the first of which did not involve a score for elaboration. The nature of the creativity score would be clearer if at least the nominal divergent thinking abilities included in the score had been stated.

It was noted by Biggs, Fitzgerald and Atkinson (1971) that their divergent thinking measure was flexibility, and that this was obtained from three versions of the Uses test, uses for a brick, a wooden box, and a tree respectively. They state that this was assessed "using category scoring" but there is no standard procedure for forming such categories. This problem will be returned to later.

The omission of scoring details is quite often acknowledged by some writers but tempered by a note that the procedures are modifications of those used by other writers, particularly Guilford and Torrance. The modifications are seldom made explicit however, and as will be seen more specifically later can involve some very significant changes. Debney (1969), studying the interrelationships of thirteen creativity tests, including Consequences and Uses, gives some general indication of the scoring procedures but not in sufficient detail for the procedures to be repeated. The Consequences Test was marked for "low grade - direct responses" and "high grade - remote consequences", but although it was noted that one mark was given for a response in the former category and three marks for the latter, how exactly to judge whether a response is 'direct' or 'remote' is not clear. The Uses Test was scored for originality according to Guilford's suggestion that a weight be assigned to each response in direct proportion to the infrequency of occurrence in the population. It is unlikely that the scoring procedure would have implemented this suggestion literally to credit every response in this way, but no details are given of any other interpretation.

Hargreaves and Bolton (1972) note that they followed the scoring procedures given in the manuals for the Minnesota Tests of Creative Thinking (Yamamoto, 1965b), and give some guidelines for the scoring of fluency, flexibility, originality and elaboration. These however were adapted and modified for flexibility and originality and the scoring for neither of these could be duplicated from the information given.

Lovell and Shields (1968), in a study of the gifted child, note that details of the divergent thinking tests used in their study and the procedures used in scoring may be obtained from the authors. This is also, of course, a possibility with other studies, and some details are in fact published with research reports, especially if published in book form (e.g. Getzels and Jackson, 1962; Wallach and Kogan, 1966; Hudson, 1968). A closer look at the details however does not reveal a consistent scoring procedure.

In the Lovell and Shield's study for example, the divergent thinking score for the Uses Test was the sum of the marks for uncommon responses and for the number of different uses. One mark was awarded for each uncommon response - i.e. not given by more than one-fifth of the population, and one mark for each different response. Other studies adopt various other criteria for uncommonness, and the interpretation of the word 'different' also varies considerably between studies. 'Different' is sometimes weakly interpreted as 'not identical', whereas Lovell and Shields apply it more strongly so that the use of a brick 'to build walls' is not credited as being different from 'to build houses'. A considerable amount of judgement is needed to apply the latter interpretation and the resulting score would seem to involve both fluency and flexibility.

When scrutinised, details of scoring procedures unfortunately show a serious lack of consistency between researchers, even though most procedures attempt to credit some or all of the abilities of fluency, flexibility and originality.

The following discussion will review some of the procedures in general use, and will recommend a precise convention and practical details for use in the present study. A final summary of the detailed scoring procedures for each test will be given at the end of the chapter. Each of the abilities of fluency, flexibility and originality will be looked at separately.

6.3 REVIEW AND PRACTICAL PROCEDURES FOR SCORING FOR FLUENCY, FLEXIBILITY AND ORIGINALITY

6.31 Fluency

Fluency in relation to tests of divergent thinking is the ability of an individual to produce a large number of responses relevant to some stimulus, verbal or figural. "It is a matter" Guilford (1967b) claims, "of the facility with which an individual retrieves items of information from his personal information in storage". This is more than a memory ability as it emphasises the ability to retrieve the information and see associations in new situations. In scoring responses for fluency he emphasises that "sheer quantity is the important consideration; quality need not be considered so long as responses are appropriate" (Guilford, 1959, 1973)

The question of how to judge 'appropriateness', or 'relevance' as it is more frequently termed, will be discussed later, in the meanwhile fluency will be defined as *the total number of relevant responses to an item.* This is the simplest of divergent thinking test scores to calculate and although an element of subjective judgement is involved in assessing the relevancy of some responses the procedure can be almost as routine as for a standardised test. Interscorer reliabilities as high as r = 1.00 have been reported for five of the Minnesota Tests including the Uses Test and the Circles Test (Yamamoto, 1962).

Unfortunately the definition of fluency is sometimes worded as the "total number of *different* relevant responses" (Torrance, 1962), and the presence of the word 'different' has confused the issue and resulted in some real discrepancies between similar sounding measures of divergent thinking.

It is evident from his Scoring Guide (Torrance, 1974) that Torrance intends the word 'different' to be interpreted as 'not identical' and he notes in the instructions for scoring the Circles Test, that "Fluency is simply the number of responses minus the number of duplications and irrelevant responses". This is made clear in a scoring example for his test of Unusual Uses (for a tin can) in which he gives credit for five similar uses of a tin, as a container for food, pencils, little toys, water and parts of things. Other experimenters assess 'difference' much more structly. Getzels and Jackson (1962), for example, observe in an appendix that in scoring their 'Uses of a brick' test, "use them to build houses; use them to build fireplaces; use them to build walls; heat them and use them as bed warmers", would be credited as only two different uses - for building and as a bed warmer. In summarising their creativity measures in the main text they note that a subject's score "depended on the number and originality of the uses he mentioned", but the details of the tests given in the appendix make it clear that they interpreted the word different in terms of 'different category of response' not simply as 'not identical'.

Getzels and Jackson comment that they originally used unnecessarily elaborate scoring procedures, but that a simplified system produced extremely similar results. In the Word Association Test for example they originally rated each script "according to the *total* number of meanings suggested; the number of different meanings, and the relative *uniqueness* of each meaning" (original emphasis). This procedure they modified to simply scoring the number of *different* meanings. They observe that Guilford (1954) used the Uses Test to assess 'ideational fluency' or 'semantic spontaneous flexibility' depending on whether it was scored for the number of responses or the number of classes into which the responses may be placed, and they note that it was included in their creativity battery "because it apparently measures the subject's ability to shift frames of reference, to use the environment in an original manner". After scrutiny it is clear that in contrast to Torrance's interpretation, Getzels and Jackson's scoring for number of different responses is, in fact, a flexibility rating.

It is suggested that such a lack of consistency in specifying and interpreting scoring procedures has been one of the reasons for the conflicting results that have occurred in research into divergent thinking. Although fluency and flexibility scores for the same test are often highly correlated no consistent patterns are likely to emerge between studies if an identical instruction to score for 'number of different responses' is interpreted by some researchers according to the examples given by Torrance and by others in the alternative manner used by Getzels and Jackson.

The publication of 'Creativity and Intelligence' by Getzels and Jackson (1962) though occasioning some detailed criticism, also sparked off an increase of research into divergent thinking often based on their procedures. The procedures referred to earlier from Lovell and Shields (1968) for example, were adapted from Getzels and Jackson and retained a strong emphasis on the need for responses to be different.

Hudson (1966) follows Getzels and Jackson's example of distinguishing between two types of child, the 'High I.Q.' and the 'High Creative', terming them 'convergers' and 'divergers' respectively. He bases the distinction on the results of two tests, Uses of Objects and Meanings of Words which he notes are almost identical to those used by Getzels and Jackson. The scoring procedures, however, were confined to quantity scoring for 'the total number of responses'. On the other hand, there is also an indication that the responses had to be different, for Hudson notes that "Nor is it clear whether 'as a container of liquids' and 'as a container for beer' should count as two uses for a barrel or only one". On balance, however, one would guess that the result is more akin to the fluency score of Torrance than that of Getzels and Jackson. This is borne out by the further details of his scoring procedures given by Hudson (1968). It is also the interpretation used by Child and Smithers (1971, 1973) who, in their investigations of convergence and divergence, note that their samples were identified in a similar manner to Hudson (1966, 1968) using "fluency scores".

Apart from considerations of relevance and duplication of responses it is likely that when specifying a score as being for fluency, experimenters mean it to be obtained by counting the total number of responses and this is the procedure adopted here. It is worth noting however that, as Wallach and Kogan (1966) point out, fluency as conceived by Guilford carried speed implications that their untimed tests tried to avoid. They still retained a score for 'total number of responses' however and although they refrained from describing it as fluency it provides an identical score if fluency is defined as suggested above.

Even in interpreting fluency in terms of total number of responses however, some assessment has to be made for the relevance of the responses.

6.311 Relevance of Responses

In any test the responses have to be related to the requirements of the task set, and divergent thinking tests are no exception. Before being credited therefore, responses have to be assessed for relevance. The requirements of a divergent test are however deliberately chosen so as not to constrain the subject, and to do him justice the scorer must not adopt a narrow view of the requirements.

It is widely suggested that teachers and others tend to favour the conformist answer and children's conventional abilities, and there is a temptation to dismiss answers that at first sight appear to be fanciful or stupid, or more generally, in opposition to the views of the scorer. Wallach and Kogan (1966) observe that responses in their study were excluded only if they "could not conceivably fit into the stimulus requirements" and they note that responses that might be considered bizarre or inappropriate were extremely rare. It is possible however that having to give their responses orally, face to face with the experimenter could inhibit the children in this respect.

Torrance (1974) cautions scorers that "any minimal response which can be reasonably interpreted and/or identified should be scored and given a point for fluency" and urges them not to give way too readily to their subjective standards for completeness. Such a caution would be well taken by a scorer who was not in sympathy with the spirit of divergent thinking assessment, though as Hudson (1968) observes "Our capacity to code (score) in ways that support our own hypotheses is well known", and some scorers could take Torrance's suggestion as a licence for 'inventive' scoring that would credit the imagination of the scorer rather than the candidate. Though the subjective element in scoring divergent tests cannot be avoided, it is important that scorers be aware of the part it plays, and attempt to subject their instinctive judgements to some form of evaluation. If more than one scorer is involved it is essential that doubtful responses are discussed and a common decision arrived at. This applies equally to judgements relating to the scoring of flexibility and originality.

Garwood (1964) suggests that any response which is 'feasible' should be credited and in practice few responses are so absurd that they have to be deleted. In accord with a further suggestion of Torrance (1974) however, fantastic or impossible responses beyond all possible reality have to be categorised as irrelevant. As a guiding principle a response should be credited if it shows a recognisable relationship to the requirements of the test.

There can however be no rigid rule for judging the relevance, appropriateness or feasibility of a response and inevitably some subjective judgements have to be made. Torrance for example allows 'throw them away' as a use for a tin can, but the writer would not.

In the Circles Test subjects occasionally draw objects within the circle with no apparent relationship to the circle itself. Unless these clearly involve another circle in their construction, when the benefit of the doubt might be given to the subject, such responses should be considered irrelevant to the requirements of the test.

It is important to realise that many divergent thinking tests, especially those used by Guilford to identify some of the factors in his Structure-of-Intellect model, have been administered and scored in a variety of ways in order to emphasise different aspects of divergent thinking. The Uses Test is the most obvious example. Guilford (1967) even describes the test differently as 'Uses for a brick' or 'Unusual uses' depending on the emphasis he wishes to convey. When he wishes to call for varied class responses he notes elsewhere that "the emphasis upon unusual uses virtually forces the examinee to go from one class of uses to another" (Guilford, 1967b). It would seem more appropriate on the other hand to suggest that the word 'unusual' emphasises originality rather than flexibility.

Following the adaptation of the tests for school use however the instructions have been more in the form of a general request to subjects to 'give as many different and unusual uses as possible for an object'. This gives an invitation for the subject to be fluent (number of responses), flexible (different responses) and original (unusual responses), and the scoring for each of these abilities is conducted independently, without considerations of difference or unusualness being applied to the judgement of fluency.

To encourage pupils to avoid the normal uses of an object the Uses Test may still be referred to as 'Unusual Uses' (Torrance, 1974) or 'Alternate Uses' (Wallach and Kogan, 1966), but there is little to indicate that the labels 'unusual' or 'alternate' in the title are taken as a specific requirement of the task when fluency is scored. Torrance however does attempt to utilize it in his scoring instructions though the net result is hardly worth the effort and he himself appears largely, in practice, to ignore it. He explains that in allocating marks for fluency the 'usual' uses of the object should be excluded. For example when the object is a tin can, he defines an acceptable use as "any relevant use other than the 'usual use' as a container for preserving food and other products. Other container responses detailed uses after the can has been emptied of its original contents are counted" (Torrance, 1974). He notes that this rather a lenient definition of unusualness and credits such replies as 'container (unspecified)', 'food container' and 'liquid container' in his list of responses. It appears therefore that the unusualness criterion is only minimally applied to the scoring for fluency, and, although it might appear in the name of the test, it seems preferable to judge the relevance of a response in terms of the general requirements of the test rather than attempt to apply an additional criterion, especially when it is later related to the scoring for originality.

Finally it needs to be noted that responses from pupils who misunderstood the requirements of the test have to be deemed irrelevant and deleted. In the present study none of the subjects completely misunderstood the requirements of the figural tests and the few responses judged to be irrelevant only minimally affected any individual's score. In the verbal tests however the misunderstanding of the directions by four candidates resulted in their total departure from the requirements of the test.

In one case 'Uses for a newspaper' resulted in a football article for a newspaper; and in the others, 'Consequences of having no hair' resulted in a story about a boy who had no hair, a list of other missing attributes such as 'no arms' or 'only one eye', and one unintelligible passage full of non sequiturs. In such cases the whole answer must be deemed irrelevant. Any estimation of the subject's score from his attempt would be invalid, though if necessary a statistical estimate could be made if sufficient information had been obtained from other similar tests. In this study the children concerned were omitted from the overall results.

This raises the question of the difference between very low scorers and those who were judged as misunderstanding the task rather than having a score of zero. In three of the cases mentioned the children were clearly on the wrong track and produced fairly competent answers for the parallel version of the test. The fourth case was more ambiguous as the pupil concerned had difficulty in writing intelligibly though her second attempt was more clearly related to the task. She also found reading difficult however and as some of the other materials had been presented in verbal form her answers to the questions were unlikely to be reliable. As a result, all four children were omitted from the overall analysis. Low scorers on the other hand were usually characterised by a short but sensible list of responses even if they exhibited limited ability to write clearly. The obvious presence of a verbal ability in answering the tests will be borne in mind when the results are interpreted.

It is important that the exclusion of certain subjects does not lead to a bias in the sample however, and such omissions need some justification. The ability of a subject to understand and follow directions may, for example, be related to his general ability and this could be a variable under discussion. This is in fact the case in the present study and the four children omitted had below average I.Q.'s. Even without these four children however a very wide range of ability is covered as the statistics of the sample will show later. With the comparatively large number of children involved therefore their omission was considered legitimate.

6.32 Flexibility

Guilford (1950) argued that flexible thinking is an attribute of creative thinkers shown by their ability to desert old ways of thinking and strike out in new directions. As discussed in Chapter 3 this view reflected earlier attempts to relate rigidity, the counterpart of flexibility, to approaches to problem solving and soon received some confirmation in a factorial study of thinking abilities (Guilford, Wilson and Christensen, 1952).

In the latter, flexibility was assessed using tests of divergent thinking by counting the number of different categories into which the subjects' responses fell. This is still the method of scoring adopted though there is considerable variation in the way in which the categories are formed.

The principle of the procedure, however, is fairly straightforward. For example, if the Uses Test asks for a list of uses for a brick, a subject who says bricks could be used for building a house, a barn, a school, a stove, a wall, a floor, and a chimney would receive seven marks for fluency but *only one* for flexibility - all the responses falling into the single category of 'building'. Another subject who says bricks could be used to build a house, throw at the cat, make a doorstop, make bookends, make a red powder, make a tombstone for a dead bird, and scratch a message, would receive seven fluency marks but also seven flexibility marks, one for each of the seven different categories of response.

Although Guilford (1959) in a similar example notes that responses of the type given by the first subject would earn "a very low score for spontaneous flexibility, because all these uses fall into the same class", he brings some slight confusion to the scoring procedure by noting elsewhere (Guilford, 1967) that a flexibility score is calculated "by counting the number of different categories of use, or, alternatively, the number of times that the examinee shifts category of use".

The latter may seem to be more exactly in line with the concept of flexibility though it is open to various interpretations which could affect its reliability and validity. Some scorers might credit a shift from one category to another even if it resulted in a return to an earlier category, though clearly this would give a very narrow concept of flexibility if the subject were to oscillate frequently between two categories. If shifts were only credited when the subject moved into a new category, then his score would simply be one less than if scored by counting each different category of response as shown above. This would give a zero score for a subject who remained in one category only and strictly speaking this would be an appropriate score for his flexibility in such a case. In practice, however, the method of 'counting each different category of response' is most commonly used, and both Torrance (1974) and Cropley (1967) recommend this procedure when giving details of their scoring procedures.

It is clearly important for experimenters to specify their method of scoring if valid comparisons are to be made between studies, and it is suggested that in the interests of consistency 'number of categories into which responses fall' should be adopted as the standard procedure for counting flexibility categories. Special note should then be made when researchers point out that their flexibility scores were found "by calculating the number of shifts amongst predetermined response categories" (e.g. Hargreaves and Bolton, 1972).

This variation in scoring for flexibility is, however, a less widespread problem than that of defining the response categories. The above recommendation for flexibility scoring, whilst clear in principle leaves open the question of how exactly the categories are constructed. Guilford leaves their construction to the scorer, and although this ensures that the categories are directly related to the population tested, it leaves open a range of options. These can vary from the construction of very general categories to others more specifically defined in terms of the predominant features of the responses.

Torrance (1966, 1974) in an effort to standardise the procedure has produced a list of categories derived from an analysis and classification of the responses of a sample of 500 subjects from kindergarten children to college students. When marking a script, the category of each response is determined from this list and entered on the scoring sheet, the flexibility score then being the total number of different categories. In cases where responses cannot be classified into any of the categories listed, Torrance allows new categories to be created but notes that this should rarely be necessary since these categories cover over 99 per cent of the responses given by the 500 subjects.

It is not surprising that most responses can be fitted into the categories as the latter are of a very general nature. As a result they are not mutually exclusive and a considerable amount of subjectivity arises in placing a response as will be shown later. What is more, the general nature of the categories means that they are only loosely related to the psychological processes that are likely to be involved in flexible thinking, and the ways in which individuals attempt the test. In particular from the experiences of the present study it appears that children of 11 years of age may approach the test in different ways to the 15-year-olds. In marking scripts it seems easier to recognise the association between consecutive responses in the answers of the younger children, even when they move into a new category. A lollipop, a balloon and a table-tennis bat, might, for example, follow each other in replies to the Circles Test, involving changes of category but little actual change of appearance in the children's drawings. Whether or not it is true that this is more likely to occur with younger children is a question which deserves further study.

It is unlikely that any general, predetermined, set of response categories could cater for such an 'associative' approach to studying children's flexible thinking, but some indication of the changes in direction of a subject's thinking can be arrived at from a set of response categories. Torrance's very general categories however, though possibly increasing reliability may be sacrificing validity. His latest scoring guide (1974) is in fact more general than the 1966 version, and although he observes that recent validity studies give confidence to the changes in scoring procedures, there are some aspects of the earlier guide which seem preferable. Some details will be considered shortly.

As noted earlier (page 196) there are some doubts about the value of scoring all three abilities, fluency, flexibility and originality on any one test of divergent thinking. As a result many researchers (e.g. Dewing, 1970; Vernon, 1971) have omitted flexibility from their investigations. As Guilford (1950) notes however - "Although there have been disappointments in the attempt to establish a common factor of this type, the concept of flexibility and of its probable opposite, rigidity, will not be downed". It was suggested in Chapter 3 that it is still an appealing concept in relation to problem solving, and the writer has shown some connection between flexibility and mathematical problems, particularly when flexibility is derived from figural tests (Richards, 1970). Flexibility scores still correlated highly with other scores from the same test, but this was less marked in the Circles Test than the Uses Test. In the former the coefficients of correlation between flexibility and the other two scores were 0.86 and 0.79, with fluency and originality respectively, while the corresponding coefficients for the Uses Test were 0.95 and 0.84. With a correlation as high as 0.95 in the Uses Test there is little variance which might distinguish the flexibility score from that of fluency, and consequently a flexibility score is included in the present study in the figural test only.

The procedures for forming the flexibility categories have yet to be established however, and these will now be looked at more closely. The practical details needed to make the procedures clear will be related to the Circles Test, although the general nature of the problems discussed is shared with other divergent thinking tests when flexibility is scored.

6.321 Practical Details of Flexibility Scoring

Although the rationale for divergent thinking abilities stems largely from Guilford's work, the establishment of the Minnesota Test of Creative Thinking (Torrance, 1962; Yamamoto, 1965b) and their subsequent commercial development by Torrance (1966, 1974) has meant that many scorers adopt the flexibility categories as listed by Torrance, and others use 'adapted versions'. Some of the drawbacks to Torrance's scheme have already been mentioned however, and as more will become apparent when the details are looked at more closely, new categories will be constructed for use in the present study. However, as most of the attempts to establish scoring procedures for divergent thinking tests has been done in relation to the Minnesota tests it would be fair to say that any new version is likely in many ways to be an adaptation of one of the Torrance versions. Whilst acknowledging this, there are a number of inconsistencies and limitations in the Torrance material that need to be clarified and amended.

In the Scoring Guide for Figural Tests, Booklet B, Torrance (1974) gives the responses of a college student to the Circles Test. Two of the responses are faces, a sleepy face labelled 'I sleepy', and a toothy face labelled 'I use Crest'. Unlike the 1966 version there is no specific 'Faces' category and the two responses are separated into categories of 'Body or body part' and 'Human beings' respectively. The justification for this separation is not clear from the scoring procedures. Other faces and human figures appear in Torrance's list of uncommon responses and these too are allocated to different categories, for example, a King's face is allocated to 'Royalty', a devil to 'Supernatural creatures' and a giant to 'Human beings'. There is a considerable amount of arbitrary judgement evident in some of the allocations.

Variations on a face are in fact common amongst responses to the Circles Test and if scored as Torrance demonstrates, could result in a subject who produced nothing but faces still getting a high mark for flexibility, and, as will be seen later, for originality. It is not uncommon for a subject to draw several faces before changing category. Some have small variations in features such as spectacles, eye patches, a hat or a moustache, others carry labels to distinguish them as a pirate, clown, school girl, princess, red Indian or sad man, while others are characters as Hitler, Edward Heath, Robin Hood or Kojak. To credit these with different flexibility categories would be appropriate for a 'faces' test, but when the responses are judged in relation to a circle it seems more appropriate to place all such responses in a 'Faces' category.

Applying this principle in general means that a response should be placed in a new category only if the scorer can recognise a definite shift of direction in the subject's thinking in relation to the requirements of the test. Although some subjective judgement still needs to be applied, it can be more closely related to the concept of flexibility than is possible when asked to fit responses into general categories such as those of Torrance.

It might appear that the resulting score for flexibility would be lower than if based on Torrance's procedures, since responses having the same basic theme, such as faces, are not put in a different category even if they contain variations on a theme. On the other hand there is no suggestion in the above principle that 99% of the responses should fit into the more specific categories compiled in the present study, and as a result the score is likely to be higher by this method.

A list of the categories is given in the Appendix. Credit for flexibility is also given when the scorer finds a response not covered by the categories listed, and often, if a number of similar responses are found, he can form a new category. Such a category need not have a collective name imposed on it if this would be a constricting influence. Instead it would be defined by the typical responses listed. Pool, pond, puddle, and swimming pool, would, for example, be likely to belong to a single category. By contrast, in

Torrance's guide 'swimming pool' is placed in a 'Recreation' category and the other responses are not listed.

Torrance's scorers, however, are encouraged to fit the responses into some category, though Torrance (1974) notes that "A response which *truly* (original emphasis) does not seem to fit into a listed category may be indicated by an 'X' and given extra credit". 'Pond' could however be put in the Geography category (which already includes 'lake') and 'puddle' and 'pool' in the 'Weather and Seasons' category. In the writer's view this would be likely to have less validity than the alternative suggested above.

Torrance (1974) observes that "in rare instances a response may not fall into any of the listed categories" and that the categories could never be exhaustive. In attempting to be as comprehensive as possible however, the categories seem to be only loosely related to the thinking process that should be reflected by the score for flexibility. As Torrance remarks "the important thing is that the testee be given credit for each different type of idea produced", but this seems hardly possible when he included in one ('Recreation') category, responses as diverse as, exercise bars, movie screen, roller coaster, swimming pool, swing, slide, skin diver, stage and strong man. Any subject giving these nine responses would appear to justify a high score for flexibility.

The introduction of very general flexibility categories for the Circles Test is one of the changes that appear in the 1974 versions of Torrance's scoring guides. In the 1966 guide for example, Torrance included categories of balls, blots, blemishes/spots, pastries/cake, hoops and optical instruments, but in 1974 these are absorbed into more general categories such as Sports, Food, Games and Science. How to score the responses of blots and blemishes/spots, is no longer clear in the 1974 guide. The earlier version though perhaps less comprehensive, is more precisely related to the actual type of response and the ways in which the pupils answer the test.

The procedure followed in the present study is more akin to this version and earlier suggestions by Torrance (Torrance, 1962). It places an emphasis on the thinking of each individual and although a number of categories are similar to those given by Torrance there is not the same attempt to force 99% of the responses into these categories. Instead it is suggested that if the defining characteristic of a response is not obviously similar to that of a listed category, the response should be awarded a mark for flexibility. The categories suggested are therefore much less comprehensive than those suggested by Torrance (1974) but it is felt that they should have more validity.

6.33 Originality

The scoring of divergent thinking tests for originality is normally based on Guilford's suggestion that it can be tested "in terms of the frequency of uncommon, yet acceptable, responses to an item" (Guilford, 1950). Although in some exercises, such as imaginative stories, originality is assessed by less statistical methods it is a practice which, as Cropley (1967) suggests, can be "both more objective and also more reliable than others".

There is considerable variation however in the way in which this principle of statistical infrequency is applied, both in the level of uncommonness at which credit is given, and in the relative weights given to different levels. Wallach and Kogan (1966) credit only unique responses - defining uniqueness literally as "of which there is but one" in the sample tested, and at the other extreme Getzels and Jackson (1962) award one mark for any response that is given by less than 20% of the sample.

Both procedures have been used by other researchers. Garwood (1964), relating personality factors to creativity in young scientists, includes an uniqueness score for responses to the Uses Test in her creativity criterion. Whereas Lovell and Shields (1968) in their creativity measure include a score for uncommon responses identical to that of Getzels and Jackson. With such a wide variation it is unlikely that the resulting creativity scores are comparable and inconsistencies are bound to arise from researches that use such differing criteria for scoring for the same nominal ability.

Each of these extremes also has its own disadvantages. Scores awarded on the basis of Wallach and Kogan's method tend to be low and therefore rather unreliable; while Getzels and Jackson's procedure gives no more credit to unique responses than to those that are given by one fifth of the population.

It is not however a straightforward case of one extreme or the other, for many experimenters use some intermediate scoring method which credits uncommon responses differentially. What is more, it is not unusual to find different criterion levels being used within a single battery of tests. Torrance (1974) for example varies both the marks awarded and the level of uncommonness credited. He allocates marks as follows:

(i) for the 'Asking' Questions test:

Frequency of a response in the population tested	<2%	2 to 4.99%	≥5%
Originality mark	2	1	0

(ii) for the Picture Construction test:

Frequency	<1%	1 to 1.99%	2 to 2.99%	3 to 3.99%	4 to 4.99%	≥5%
Originality Mark	5	4	3	2	1	0

(iii) for the Circles test:

Frequency	<2%	2 - 4%	5 - 9%	≥10%
Originality Mark	3	2	1	0

Torrance does not justify his use of different levels of frequency and their corresponding originality marks, though it is possible that these are related to differing patterns in the distributions of responses for different tests. On most of his verbal tests the marks are the same as that for the Asking Questions test though the frequency percentages credited are not stated explicitly. In an earlier publication Torrance (1962) after giving weights for different frequencies notes that "More simply, an originality score may also be determined by counting the number of responses not in the zero category". This alternative is not suggested in the scoring guides for the later versions of his tests.

Haddon and Lytton (1968), using five tests from Torrance's battery, reflect some of the uncertainty in the 1962 versions of the scoring procedures, and vary their scoring from giving one point to any uncommon response (Circles Test) to a scale of 1 to 4 marks for degrees of unusualness up to a cut-off point of about 5% (in the Uses and Common Problems Tests).

A cut-off point of 5% is favoured by a number of researchers. Dewing (1970) adopts a 5% cut-off point, and the same weightings as Torrance for the Asking Questions test, and is also consistent in applying the same criterion for both of the tests she uses. This is particularly important when a total creativity score is to be calculated. Barker Lunn (1970) also adopts this procedure for assessing originality and for each of the divergent thinking tests used in her study.

Cropley (1967) recommends awarding points for originality up to a maximum of 4 for a unique response, with a cut-off point at 15%. He suggests the following mark allocation:

Frequency	<1%	1 to 2%	3 to 6%	7 to 15%	>15%
Originality Mark	4	3	2	1	0

Vernon (1971) adopts the 15% cut-off point as suggested by Cropley but with the following simplified weighting categories:

Frequency of response in the sample tested (n = 100)	<5	6 to 14	≥15
Originality Mark	2	1	0

He also suggests constructing the list of uncommon responses from a random sample rather than the total population of scripts. This procedure reduces some of the labour involved in listing and weighting every response, and will be returned to later.

Cropley (1967) notes that his originality categories, expressed in percentages, are apparently uneven, the one point category covering a range from 7 to 15 per cent, the next only from 3 to 6 per cent and so on. He explains this by noting that the percentage limits for each category correspond to equal standard-score distances along the x-axis of a normal curve - "Thus, the distance between 7 and 15 per cent is about 0.44 standard deviations; as is the distance from 3 per cent to 6 per cent and so on". It is not made clear why equal standard deviation intervals were considered preferable to equal percentage intervals, though the effect is to add more weight to the 'original' end of the scale than equal percentage intervals over the same range would have done. This is achieved in the present study by increasing the weighting of unique responses from an otherwise linear scale of marks based on equal percentage intervals.

Some confusion could arise from the way in which Cropley presents his intervals. In ending intervals at 2% and 6% and beginning the next intervals at 3% and 7% respectively, he loses the two 1% transition intervals and it is not immediately clear in which intervals intermediate scores should be placed. This could cause some difficulty in trying to apply the scheme to a sample whose size was not 100. What mark, for example, should be given to a response given by 7 out of 120 subjects, i.e. by 6.7% of the population? It could be taken as 7% 'to the nearest whole number' and awarded 1 mark, or alternatively as occurring in less than 7% of the responses be given 2 marks.

Calculating the end points of Cropley's intervals in terms of standard deviation units of the normal curve helps to clarify the intervals intended:-

Percentage frequency contained in tail of the normal curve	1	2	3	6	7	15
Deviate in standard score units	2.33	2.05	1.88	1.55	1.48	1.04

The sizes of the intervals from 1 to 3%, 3 to 7%, and 7 to 15% are 0.45, 0.40 and 0.44 s.d. units respectively and clearly these are the intended intervals. (The interval from 0 to 1% is indeterminate.) The actual end points, however, are still nominally in two categories. A clearer procedure would be to note the cut-off points for each interval and whether or not it is included in the interval. Torrance does this for two of the distributions given earlier by noting that the intervals end at 1.99% etc. though he also gives an indefinite version for the Circles Test.

Many researchers begin their distribution with the award of maximum originality marks for responses given by less than 1% of the population, though when the size of the population tested is less than 100, no responses can occur in this category. When this happens it is suggested that the first category should be that of an unique response whatever its percentage frequency.

6.331 <u>Use of a Sample of Scripts for Listing Uncommon Responses</u>

Instead of forming the list of uncommon responses for the whole population of scripts, Vernon (1971) suggests that basing the procedure on a random sample of 100 scripts would be less time consuming and still adequately reliable. A similar system has also been used by other experimenters including the writer (Richards, 1970). The scoring guides produced by Torrance (1974) include lists of uncommon responses based on the same sample of 500 subjects of all ages from kindergarten to college, which he used for establishing flexibility categories. The representativeness of the sample, however, is even more questionable for originality than it was for flexibility. Responses which are original for a group of 7-year-olds may well be commonplace for a group of college students and it is also unlikely that his distribution would be suitable for populations outside the U.S.A. As Cropley (1967) argues, originality distributions should be compiled for each group of subjects tested, such as year-group or school class. This is the practice followed in the present study, with a sample of the group being tested.

The size of the sample necessary depends on the composition of the group, the more homogeneous it is the smaller the sample need be. Whatever the sample size however the list of responses can never be guaranteed to be exhaustive. In observing this in her study, Barker Lunn (1970) goes on to note immediately that "A certain amount of discretion had to be employed in interpreting responses, and where queries arose these were discussed by markers and members of the research team". This is an observation often made on the scoring of divergent thinking tests in general, but should not be taken as specific advice on what to do with responses not appearing in the originality list. A more reliable procedure is to score any new responses as original even if they are repeated in the scripts not comprising the distribution sample. The writer has found this a reasonable procedure for verbal tests, only a limited number of new responses actually occurring. In the Circles Test with many more responses however, the use of a sample is likely to be less exhaustive and the procedure consequently less valid. Vernon (1971) acknowledging that complications such as this are bound to arise in any sampling procedure where the frequency of occurrence of responses in the sample is unlikely to be identical to that in the whole population, nevertheless estimates that "the resulting error probably has little effect on total test scores". In the present study a sample of scripts will be used for scoring the verbal tests, but in the interests of validity the whole population will be used to draw up the frequency distribution for the Circles Test.

The basis of the sampling procedure outlined by Vernon is that the sample should be a random one taken directly from the population being tested. This is not the case when using a 'standard' list of uncommon responses such as those compiled by Torrance, and the problem of how to deal with responses not on the list is then a rather different one. Torrance (1974) suggests that although most new responses will justify the maximum originality mark the scorer should exercise his judgement of the "creative strength" of any new response. Torrance's guidance on how to assess creative strength is based on the scorer acquiring experience and familiarity with the appropriate originality lists in the scoring guide. In this way he believes the scorer should gain a feeling for the difference between non-scoring and scoring responses. The former, 'obvious' responses are he says, characterised by requiring little intellectual energy, while in contrast the scoring responses require more intellectual energy to be "beyond what is learned, practical, habitual and away from the obvious and commonplace".

It is true that familiarity with a large number of scripts does give one a feeling for the clearness and originality of responses and that in a non-standardised test some discretion has to be applied. In the interests of interscorer reliability however the scoring procedures should attempt to keep this aspect to a minimum. Vernon (1971) expresses the view that it is reasonable to allow the scorer to use some subjective evaluations of the value, cleverness or originality of unusual responses, provided they are recorded and applied consistently. The latter is particularly difficult if more than one scorer is involved in the marking. Ideally the whole scoring, including the construction of the frequency distribution should be done by one person. This principle has been adhered to in this study, all the marking except for certain routine tasks carried out under his close supervision, being done by the writer.

6.332 Summary

The preceding discussion of the scoring for originality has attempted to illustrate some of the differing interpretations of the criterion of uncommonness, and some of the accompanying problems including the use of a 'standard' scoring guide for awarding originality marks, and a sampling method for determining one's own frequency distribution. As a result the writer has arrived at the following principles for the marking to be used in the present study:

1. To reject the use of a 'standard' list of uncommon responses from a scoring manual such as that of Torrance (1974).
2. To construct frequency distributions of uncommon responses for each test used and for each of the two year groups tested.
3. To adopt a scale for awarding originality marks by weighting the responses in similar way to that given by Cropley (1967) but based on the percentage intervals as shown:

Frequency	*f≤1%	1%<f≤5%	5%<f≤10%	10%<f≤15%	f>15%
Originality Mark	5	3	2	1	0

* or f = 1 (unique) if n<100

(This allocation is applied to all the tests apart from those which had to be scored according to earlier procedures for the sake of direct comparison.)

4. To use a sample of scripts to form the frequency distributions for the verbal tests, and the whole population for the figural tests.

6.333 <u>Practical Details of Originality Scoring</u>

A number of practical details have now to be considered if the above principles are to be put into practice reliably. The level of interscorer reliability places its own limits on the degree of reliability and validity that can be in relation to other variables, and scorer reliabilities will be calculated for each of the tests used in this study.

Practical problems in scoring for originality become apparent as soon as one begins to construct one's own list of uncommon responses, or attempts to understand the principles involved in the examples and frequency distributions prepared by other scorers. It is essential, as Torrance (1974) notes, for prospective scorers to gain familiarity with the construction of the originality lists if they are to apply the same criteria to any new responses that occur. In scrutinising his procedures however, a number of practical issues emerge that need resolving before similar procedures can be established for use in the present study.

As an example, Torrance, for the Circles Test, gives originality marks to such responses as a king's face, a toothy face; a billiard ball, a golf ball, a tennis ball, a basket ball; Venus, Pluto and Mars. In the present study these will be collectively recorded as 'faces', 'balls' and 'planets' respectively, and given originality marks according to the frequency of the group rather than that of an individual response, *unless the latter possesses some special defining characteristic that identifies it in different relation to the requirements of the test than other members of the group.* This principle, similar to that adopted for judging flexibility, can also be applied within flexibility categories and is adopted for the scoring of all the tests used.

In the Circles Test it means that most of the small variations to 'faces' referred to earlier such as moustaches, spectacles or eye patches; or the addition of different labels such as pirate, king, clown or princess are not likely to qualify for originality; nor are the labels 'golf', 'billiard', or 'basket' if used to describe a ball. In a similar way no more credit is given for the specification of a circle as 'Mars' than for the general description of the circle as 'a planet'. The defining property of these examples in relation to the circle is basically that of a 'face', 'ball' and 'planet' respectively, and it seems unreasonable that in Torrance's lists 'Mars' should receive maximum originality marks when 'a planet' as a common response, receives none. On the other hand if the circle itself formed the rings of Saturn, or a diver's helmet, or a ball and chain, then the response would qualify for separate originality marks.

In the verbal tests some general responses are quickly established, and once again any slight variations or specific examples are not credited separately from the general response. An extension to a commonplace answer, however, can be a definite source of originality in the verbal tests in a way which is not possible in the figural tests. The ability to make an original verbal response often appears to be more intellectually demanding than making an uncommon figural response, though the latter may encourage a more unconstrained imaginative ability. This question will be discussed when the interrelationships between the tests are looked at later. The following examples illustrate the details in applying the principle for awarding originality marks for the verbal tests.

In the Consequences Test, for example, a common theme when responding to the consequences of having no hair is 'no need for articles to make it look tidy'. This is often expressed in terms of "no need for brushes, combs, etc." (sic). The use of 'etcetera' reinforces the writer's view that a mention of specific items does not add to the quality of response, and should not form separate items in tabulating the uncommon replies. The specific mention of brushes, combs, rollers, curlers or hairgrips are consequently all included in terms of the above general theme. A similar example occurs when subjects note that grooming accessories such as hair cream, hair spray or shampoo would not be necessary. 'Hair restorer' however has a different relationship to 'having no hair' than the other applications and it could be argued that it deserves separate credit.

As a consequence of 'no need to eat' a common category of response is 'no need to keep animals for food' and this is scored as a collective response to include specific responses such as 'no need to keep cows, sheep or pigs'. 'No need to keep vegetables' is a similar example, though an extension of the idea to the response 'only flowers would be grown in gardens' is judged to be worthy of separate credit.

In the Uses Test a common response to being asked for uses of a spoon is 'to eat with' and this is taken to include specific responses as 'to eat food', 'to eat pudding' or 'as a desert spoon'. 'To dig with' is another quite common theme though it is less specific in relation to using a spoon than 'to eat with'. The addition of more specific information such as 'to make sandcastles' or 'to plant things with' can consequently result in ideas that deserve separate credit for originality even though they are in the same general category of 'digging with'.

In responding to Uses for a newspaper, the suggestions that it be used 'to carry vegetables', 'carry potatoes' or 'carry things' do not seem to differ sufficiently to deserve separate credit for originality though to 'carry water' is rather different. Similarly to use a newspaper 'round the rim of a bin to keep it airtight' is judged as being a very different response to 'use as a bin liner'.

In tabulating responses for originality some subjective element is inevitably present though the principle discussed above is not usually difficult to apply. Cropley (1967) claims that tabulating the frequency of every response in an objective way of scoring for originality, but this seems an over-optimistic statement in view of the practical problems discussed above. Although it is theoretically possible to tabulate every single response, giving separate credit for every different statement including those that appear to be different ways of saying the same thing, such a method would distort and invalidate the notion of originality on which the scoring is based. Even if some judgements can be avoided by using predetermined lists such as those of Torrance, judgements of new responses have to be made and it is suggested that the above procedures are likely to be more valid than an attempt to apply those implicit in Torrance's lists.

One further problem needs to be clarified before the procedures can be put into operation. In the Circles Test, for example, Torrance lists both Venus and Mars as deserving maximum marks for originality but gives no indication of a different procedure when a subject gives both responses. It is difficult to justify the award of two lots of originality marks one for each of such responses. It is also difficult to apply a rigid rule as the degree of similarity varies enormously. Although it seems reasonable to credit for originality only one of two responses as similar as Venus and Mars, it is more difficult to judge the award of marks to less similar responses.

If the responses are in different flexibility categories it usually follows that they should have individual credit for originality. It is when they are in the same category but yet show some variation that judgements are most difficult to make. For example, in the Circles Test, the writer credits one pupil in the present study with originality marks for both 'volcano' and 'moon crater', though it could be argued otherwise. It is not necessarily an all or nothing situation and it is possible to allocate reduced marks to the second of two such responses, depending on their degree of similarity, but this introduces another source of unreliability and a more standard procedure is recommended in this study. If two similar responses are given, both individually deserving credit for originality, the higher scoring response only should be scored. When more than two related but not identical responses are given it is suggested as a general guide that credit for originality should not be given for more than TWO of them. If, for example, bomb, mine, hand grenade, and torpedo are all given by one pupil it is suggested that only the two best responses be awarded credit. In this study these responses would qualify for originality marks of 2, 3, 3 and 5 respectively, but if given together would therefore be awarded a total of eight marks.

In assessing the originality of responses to the Consequences Test (for no need to eat) it is common to find a response such as 'no need to clean teeth' being developed into additional responses such as 'no need for toothbrush' or 'no need for toothpaste'. A similar guideline to the above can be applied whereby extensions on the same theme are given separate credit for originality. It is suggested that the main idea plus one extension, i.e. a total of TWO only are credited. In this study the originality marks for these responses would then total three, one mark for 'no need to clean teeth' and two marks for 'no need for toothpaste'. Similarly 'no need for farms, no need for a tractor, no need for a trailer' would be marked for the theme and one extension. In this case 'no need for farms' which is a common response would score nought and 'no need for a tractor' two, giving a total score of two.

Similar considerations are applied to the scoring of the Uses Test. For example two responses observing that a newspaper could be used 'as a lining for drawers' and 'as a lining for a dog's basket' are sufficiently different to receive separate credit even if given by the same person, though the addition of 'as a lining for an animal's cage' or 'as a lining for cupboards' would not appear to justify further originality marks for the same person.

6.4 A NOTE ON THE TIME INVOLVED IN DIVERGENT TESTS

It must be pointed out that the scoring procedures adopted in this study though using a sample of scripts to form the frequency distributions for originality in the verbal tests, are still enormously time-consuming. For a study of this type however it is essential to base the scoring methods on the most established of previous procedures, with enough detail to create a firm basis from which any modifications can then be recommended.

Few researchers acknowledge the time that is involved in scoring for anything more than fluency, though Vernon (1971) gives some details which put the task in perspective. Commenting on the marking, with three fellow scorers, of the scripts of about 400 pupils, each of whom spent 140 minutes on divergent thinking tests, he notes that "All four of us were largely engaged in this for almost six months".

Although it is not clear how exactly to interpret the word 'largely' some estimates might make the information more enlightening. If the scorers each spent three days per week marking it means that about 300 scorer-days were needed to mark just over 900 testee-hours of work. This puts the time needed at about one-third of a day or, say, two hours, for each hour of candidate's work.

Some further details can be provided from this study, basing the time taken on those periods when nothing else but scoring took place. In practice it is difficult to find long periods of time to devote exclusively to scoring, though an interruption for any length of time is bound to affect the efficiency and distort the subjective element of the marking. It is thus recommended that any one section of the marking be done in as continuous a period as possible. In any one day, however, it is difficult to maintain one's efficiency after about six or seven hours of marking.

Scoring for originality is the most time-consuming element in the procedures as frequency distributions have to be drawn up not only for each test but also for each item in the test. Separate distributions were also drawn up for each age group tested in this study as proposed earlier. As a result the Uses Test alone necessitated five separate originality distributions, two for the 11-year-old population and three for the 15-year-olds.

Using a sample of scripts reduces some of the work, though even so the construction and organisation of a single distribution usually took the writer between 10 and 15 hours for a sample of 42 scripts in the verbal tests. For larger samples, particularly the whole populations of 165 and 139 in the case of the Circles Test the time taken was much greater.

For the Circles and Squares Tests it was found convenient to translate each completed script into a line of symbols. This releases the scorer from the need to go through the original scripts for a second time and means that decisions about the identity of a response have only to be made once. This eliminates the chance that on a second reading an obscure response could be given a different interpretation. Some common objects are quickly recognised, and these can simply be recorded as a tick or a figure '1'. Responses in the same flexibility category as previous ones are enclosed in brackets. Objects that might attract originality marks are listed by name. The latter are also enclosed in brackets if in the same category as previous responses. The frequency distribution of uncommon responses is then drawn up from these representations of the candidates scripts and originality marks allocated to each response according to its frequency. Finally the subject's score is obtained by counting the number of responses, for fluency; the number of unbracketed responses, for flexibility; and the weightings of each named response, for originality. This procedure for the Circles Test with 165 subjects, itself took a total of 10 days.

In all, the scoring of the divergent tests, excluding the time spent on tests scored during previous work, needed about 320 hours or approximately 53 days of marking time. Over 300 pupils were involved, and in total 287 testee-hours of work was marked. Thus just over one hour of scoring time was spent for each hour of testee time - as compared to the estimate of about two hours in Vernon's study.

In the latter no flexibility scores were calculated, though frequency lists had to be tabulated for 30 different items. In the present study fewer (12) frequency distributions had to be constructed, six for each of the populations tested, namely 2 for Circles, 1 for Squares, 5 for Uses and 4 for Consequences. (One complete frequency distribution for each test is included for reference in the Appendix.) Vernon also used a larger sample of scripts to draw up his distributions though it is unlikely that his total number of responses was greater for any item, as he allowed much less time per item than that available in the present study.

Notwithstanding the differences it is evident that when originality lists have to be compiled, the scoring of divergent thinking tests takes an enormous amount of time. Cropley (1967) arguing for the use of an originality score in spite of the two perusals of scripts which are then necessary, notes that "the method is time-consuming but, unfortunately, when one is dealing with an essentially non-mechanical process like creative thinking, mechanical methods of scoring are of little use". Hargreaves and Bolton (1972) come to a different conclusion however and suggest that the effort needed to score for anything more than fluency does not justify the small amount of extra information gained, and that "to calculate fluency scores only would render divergent tests much more amenable to automated scoring and eventually result in their more widespread use". They are certainly correct in their latter observation, though it is by no means clear what the cost would be to the validity of divergent thinking tests.

What is clear, however, is that there is a great need for some simple scoring procedures if they can be shown to be valid. Considerable attention has already been given to clarifying the scoring procedures used in this study and the patterns of their inter-relationships and their relationships with other measures will be investigated later. The writer should then be in a position to make some recommendations on the above question.

Although often tedious as well as time-consuming it is right to acknowledge that there are also times when scoring divergent tests can be stimulating and amusing. The responses are often clever and humorous, and the consistency with which certain subjects produce such responses is encouraging subjective evidence of the validity of a divergent dimension. How consistent is any such trait and to what extent the quality of a subject's response is related to his fluency, are further questions to be investigated later.

6.5 SYNOPSIS OF SCORING DETAILS FOR THE DIVERGENT THINKING TESTS

6.51 Circles Test

(i) Fluency

One mark was awarded for each relevant response.

(ii) Flexibility

One mark was awarded for each different category of response. A list of categories was constructed from an analysis and classification of the responses of all the subjects in the present study. The list of categories is included in the Appendix.

(iii) Originality

Marks were awarded to a response in relation to its frequency of occurrence in the responses of the sample tested. Responses given by more than 15% of the sample were regarded as common and given no credit. Separate frequency distributions were drawn up for each of the two populations tested. The complete distribution for the 11-year-old population is given in the Appendix.

The following scales were adopted for awarding originality marks:

(1) 11-year-olds n = 165

Frequency (per cent)	f≤1%	1%<f≤5%	5%<f≤10%	10%<f≤15%	f>15%
Frequency (number of responses)	1 (unique)	2-8	9-16	17-24	≥25
Originality Mark	5	3	2	1	0

(2) 15-year-olds n = 139

Frequency (per cent)	f≤1%	1%<f≤2%	2%<f≤4%	4%<f≤7%	f>7%
Frequency (number of responses)	1 (unique)	2-3	4-6	7-10	≥11
Originality Mark	6	3	2	1	0

NOTE:

(i) The allocation of originality marks is based on different frequencies in these two samples. As discussed earlier, the writer now favours that used with the 11-year-old sample and this has been used consistently in all the creativity tests used with that sample. The allocation for the 15-year-olds was dictated by the need to compare the results with those of a previous study.

(ii) In each of the above cases the frequency distribution was drawn up using all the data available, that is using the responses of the whole sample, the numbers being 165 and 139 for the 11-year-old and 15-year-old samples respectively.

(iii) The frequency distributions reproduced for reference in the Appendix are those arrived at with the main 11-year-old population, using the present recommended procedures.

6.52 Squares Test

This is an alternate form of the Circles Test used for assessing test-retest reliability with the 11-year-old population. It was scored in exactly the same way as the Circles Test. The flexibility categories showed a very substantial overlap with those drawn up for the Circles Test and a joint set of categories was formed as given in the Appendix.

The frequency distribution for originality was based on the whole sample of 65 scripts used in the reliability assessment and is also included in the Appendix.

The originality marks were allocated as shown:

11-year-old sample n = 65

Frequency (per cent)	f≤2%	2%<f≤5%	5%<f≤10%	10%<f≤15%	f>15%
Frequency (number of responses)	1 (unique)	2,3	4,5,6	7,8,9	≥10
Originality Mark	5	3	2	1	0

6.53 Uses Test

(i) Fluency

One mark was awarded for each relevant response.

(ii) Originality

Marks were awarded for uncommon responses on the basis of statistical infrequency. To allocate the appropriate mark separate frequency distributions were drawn up for each item in the tests, and on a random sample of the population tested. With the 11-year-old population only single items were used, 'a newspaper' in the initial testing and 'a spoon' in the retesting. In the 15-year-old population three items, newspaper, spoon, and string were included in the same test and the subject's marks on each item were added to give his total score on the test.

Originality marks were awarded as shown below, and a complete frequency distribution for Uses of a newspaper, based on the recommended 11-year-old frequencies is given for reference in the Appendix.

(1) 11-year-olds (25% sample) n = 42

Frequency (per cent)	f≤3%	3%<f≤5%	5%<f≤10%	10%<f≤15%	f>15%
Frequency (number of responses)	1 (unique)	2	3,4	5,6	≥7
Originality Mark	5	3	2	1	0

(2) 15-year-olds (38% sample - as in original study) n = 53

Frequency (per cent)	f≤2%	2%<f≤4%	4%<f≤10%	10%<f≤21%	f>21%
Frequency (number of responses)	1 (unique)	2	3-5	6-11	≥12
Originality Mark	5	3	2	1	0

6.54 Consequences Test

(i) Fluency

One mark was awarded for each relevant response.

(ii) Originality

As in the other tests the originality marks were awarded for uncommon responses in relation to their statistical infrequency. Separate distributions were drawn up for each item, two for the 11-year-old population and two for the 15-year-olds. In the case of the 15-year-old sample both items were included as one test and the subject's marks on each item were added to give his total score. The frequency distribution for Consequences of having no hair, arrived at with the 11-year-old sample is reproduced for reference in the Appendix.

The originality marks were allocated as shown below.

(1) 11-year-olds (25% sample) n = 42

(As for the Uses Test, (1) above.)

(2) 15-year-olds (28.5% sample) n = 40

Frequency of a response in the population tested				
	percentage cut-off point	3%	20%	20%
	number range	1 (unique)	2-8	9
Originality Mark		2	1	0

6.55 Selection of the Distribution Sample for Assessing Uncommon Responses

Whenever a sample of scripts was used to construct the frequency distribution of uncommon responses, the scripts were chosen in a stratified random manner from the total population. The random selection was stratified to the extent that it took place within each class and for boys and girls separately, so that each school class and each sex contributed the same proportion of pupils to the sample as it did to the whole population. Within these constraints the pupils were selected using a table of random numbers (Yamane, 1967).

6.56 Examples of Common and Uncommon Responses to Each of the Divergent Thinking Tests

(Selected from the 11-year-old population.)

Circles Test

Examples of common responses were: faces, wheels, fruits, tennis racquets, and roadsigns. Uncommon responses included: a g rm, a corn plaster, a link in a chain, a trout hatchery and a finger-print.

Squares Test

Examples of common responses were: books, boxes, radios, pictures and windows. Uncommon responses included: a bale of hay, a drain cover, a piano, a playground, a sausage on tomorrow's world.

Consequences

(i) No need to eat

Examples of common responses were: no need for cutlery, no foodshops, wouldn't get fat, we would have money, no need to grow vegetables. Uncommon responses included: in space there would be no problem about food floating about, there would be more space to build houses, no harvest, no custard pies to throw at clowns, Fanny Craddock would be out of a job.

(ii) No hair

Examples of common responses were: no need for combs/ hairbrushes, we would all look like Kojak, our head would get cold, barbers would be out of a job, everyone would buy a wig. Uncommon responses included: it would be easier for artists to paint people, the R.A.F. wouldn't be known as the 'Bryl-cream boys', lollipops would be in demand (an oblique reference to Kojak), toys would not have read hair, wouldn't have the trouble of saying "Your hair looks nice", Indians wouldn't have names like 'Flowing hair' or 'Tuffty hair'.

<u>Uses</u>

(i) Newspaper

Examples of common responses were: to read, cover table when painting, sit on, use as a blanket, make paper boats/ aeroplanes, make papier maché. Uncommon responses included: make a tent, hold roses to keep thorns off, growing mustard and cress, holding to be seen at night on the road, hiding nasty sights like dead bodies.

(ii) Spoon

Examples of common responses were: to eat with, dig with, take medicine from, egg and spoon race, flick things, play the spoons. Uncommon responses included: a candlestick, a paddle for a boat, bolt a door, for a collection of old spoons, signal with it in the sun, fix to shoes for tap dancing.

CHAPTER SEVEN

INTERSCORER RELIABILITY

7.1 INTRODUCTION

Whenever scorers can use a certain amount of their own judgement in scoring tests, it is important to know the extent of the agreement between different scorers. This is usually reported in terms of the coefficient of correlation between the scores awarded to the same sets of scripts by two independent scorers. Although such a coefficient of interscorer reliability does not, of itself, indicate that a test is reliable, it is a necessary part of any overall picture of a test's reliability.

Reports of interscorer reliability for tests of divergent thinking suggest that high levels of reliability can be achieved by independent scorers provided they are familiar with and understand the criteria for scoring. Anastasi (1968), reviewing the tests used by Guilford and Torrance in their investigations of divergent thinking, suggests that "With a reasonable amount of practice in following the instructions and studying the examples in the manuals, such tests can probably be scored with satisfactory consistency, although the scoring process is quite laborious and time consuming". She points out that although scorer reliability is not reported by Guilford in the manuals, other investigators who have employed the tests in research have obtained coefficients of around 0.90. Reports of interscorer reliability are not common however, most researchers choosing to assume that the reliabilities are high, usually as a result of claims made by Torrance in his early publications and the latest issue of his tests (Torrance, 1962, 1974).

An emphasis on the need for scorers to study the scoring guides thoroughly, however, is noted by Torrance. In attempting to widen the scope of his tests for use in research and by the educational community at large, he acknowledges that some concern was felt about "whether reliable scoring could be achieved by classroom teachers by studying the scoring guides". Presenting the results of an experiment in which classroom teachers untrained in the use of the Torrance Tests of Creative Thinking scored a sample of 25 test booklets, completed by children at the grade level at which they were teaching, Torrance notes that when compared to the results obtained by a trained scorer, mean interscorer reliabilities of 0.96 for fluency, 0.94 for flexibility and 0.88 for originality were obtained on the figural tests. Mean reliabilities for the verbal tests were 0.99 for fluency, 0.97 for flexibility and 0.94 for originality. A comparison of the individual scores however is more illuminating than the averages. For the verbal tests only one out of 15 coefficients of reliability was below 0.90, though for the figural tests, with a different number of teachers, nineteen out of 28 coefficients exceeded 0.90, eight lay between 0.80 and 0.89, and one equalled 0.66. Five of the seven reliability coefficients for originality were below 0.90. Torrance points out that low interscorer reliabilities appeared to result most often from failure of the scorers to scan adequately the originality weights listed in the scoring guide.

In the above experiment the teachers had been asked to follow the scoring guides as carefully as possible but no specific training was given. Scorer reliability can be increased, Torrance suggests, by employing training in which scoring rationales are discussed, practice is given in scoring a set of four or five booklets, and the resulting scores discussed with an experienced scorer. He claims that this procedure generally results in almost no differences in mean scores between markers, and coefficients of reliability in excess of 0.90. If differences in means are significant at the 0.10 level or less, or if there are intercorrelations of less than 0.90 between composite scores for any of the abilities tested, he recommends that further training should be provided, though this, he says, is rarely necessary. In his own research he notes that scorers are excluded if they cannot maintain reliability coefficients of above 0.90 or if their mean scores show significant differences from those of an experienced scorer.

Comparing the scores awarded by an experienced scorer with those given by a scorer undergoing training, Torrance quotes the following reliabilities (Table 4), for a sample of 100 verbal and 100 figural booklets. He does not however give any details of the sample, and this is an important omission as the age of the subjects is likely to affect the type and quality of their answers and present different problems to the scorer.

	Verbal booklet			Figural booklet			
	Flu.	Flex.	Orig.	Flu.	Flex.	Orig.	Elab.
Coefficient of reliability	0.99	0.95	0.98	0.98	0.98	0.86	0.92

(There were no significant differences in mean scores n = 100)

Table 4 Interscorer reliabilities between an experienced and an inexperienced scorer for the Torrance Tests of Creative Thinking (Torrance, 1974)

The only coefficient not reaching Torrance's self imposed level of 0.90 is that for originality on the figural booklet, and he notes that the tendency for originality to have lower scorer reliability than the other scoring categories is something to which special attention might be given in the training procedures. Notwithstanding this, the level of scorer reliability is consistently high and justifies, for Torrance, the inclusion of the above measures in his scoring procedures. Scoring for elaboration on his verbal tests is already excluded by Torrance, since earlier information suggested that scorer reliability for elaboration is not high enough for the procedure to be recommended.

In general the high levels of interscorer reliability cited by Torrance are very encouraging, though it needs to be noted that the coefficients for each ability are calculated between scores for the booklets as a whole, that is between the composite scores of seven activities in the case of the verbal booklet and three for the figural booklet. The use of composite scores has the effect of smoothing out individual test differences and increasing the overall reliability coefficient. This is a valid procedure for reporting information regarding the interscorer reliability of a testing battery designed to yield composite scores, though the coefficients are likely to be exaggerated estimates of the scorer reliabilities of individual tests. In view of the widespread use of the individual sub-tests, details of their separate interscorer reliabilities may be more widely applicable. In the present study interscorer reliabilities will be reported separately for each of the divergent thinking tests.

Yamamoto (1962) reports such reliabilities for a number of the sub-tests of Torrance's Minnesota battery including Unusual Uses and Circles. They were calculated from 64 sets of scripts marked by two independent scorers and resulted in the following coefficients: Circles Test - Fluency 1.00, Flexibility 0.91, Originality 0.98; Uses Test (toy dog) - Fluency 1.00, Flexibility 0.84, Originality 0.92; Uses Test (tin cans) - Fluency 1.00, Flexibility 0.87, Originality 0.98.

The scripts, however, were obtained from a sample of school teachers and would not have given the same problems of legibility and relevance that arise from the scripts of young children. The reliability coefficients and therefore higher than might be expected from a population of school children.

In a further study, involving 5th grade children of 10 and 11 years of age, however, Yamamoto (1965c) gives inter-scorer reliabilities ranging from 0.97 to 0.99 for a number of independent scorers on samples of 64 and 76 scripts. The reliabilities in this case are not strictly comparable with the above, being obtained for a single index of 'creativity' obtained from summing scores for fluency, flexibility, originality, adequacy, and elaboration from five of Torrance's tests.

Torrance's tests were also used by Treffinger, Speedie and Bruner (1974) in another investigation of 5th grade children's creativity, and the authors report that "acceptably high levels of interscorer reliability (all greater than 0.90) were obtained". The tests were marked by trained scorers and yielded six scores consisting of verbal and figural measures of fluency, flexibility and originality. Each score was contributed to by a number of sub-tests.

Coefficients of interscorer reliability are not reported by Wallach and Kogan (1966) for the tests in their study of children's thinking, but they note that in judging uniqueness there was usually 98 or 99 per cent agreement on the responses to each item. Disagreement between scorers on only 1 or 2 responses in every 100 is indicative of a very high level of consistency in their judgements of the acceptability of unique items.

Interscorer reliabilities for each of the tests used in the present study were computed by Wodtke (1963) in an investigation of the reliability and validity of the Torrance tests. On the basis of the scoring manual provided by Yamamoto (1962), the scores of two independent markers on a sample of 66 fourth and fifth grade pupils were compared. Although the I.Q. of the sample is not specified, Wodtke describes it as an "unrestricted sample of elementary school pupils". The tests, age range, and ability in Wodtke's experiment are therefore all similar to those in the present study. One of the scorers was inexperienced and one had some previous experience. Wodtke found reliability coefficients for the individual tests ranging from 0.82 to 0.99. Details for the Uses, Consequences and Circles Tests were as shown in Table 5.

N = 66

	Unusual Uses	Consequences	Circles
Fluency	0.99	0.98	0.98
Flexibility	0.85	0.96	0.96
Originality	0.98	-	0.99

Table 5 Coefficients of Interscorer Reliability between two independent scorers for sub-tests of the Torrance Creativity Battery (Wodtke, 1963)

These compare favourably with the composite coefficients reported earlier from Torrance. In general the evidence suggests that a high level of scorer reliability, reflecting a good degree of objectivity and agreement, can be achieved on divergent thinking tests by relatively inexperienced scorers. This has been achieved by employing scoring procedures for fluency, flexibility and originality. Divergent thinking tests which are not amenable to such procedures however have had to adopt less reliable forms of scoring.

One of these, 'Imaginative Stories' (Torrance, 1962), though still frequently used as a measure or criterion of divergent ability has now been omitted from Torrance's battery. With detailed procedures for scoring this test, Yamamoto (1961) could only report interscorer reliabilities of 0.79, 0.80, and 0.76 for a number of independent scorers; and for the same test Goldman and Clarke (1967) and Torrance (1962) report scorer reliabilities as low as 0.72 and 0.78 respectively. These are higher than have been reported for the marking of English compositions in the G.C.E. examination (Schools Council, 1966), but not as high as experimental marking procedures reported, in the same publication, based on a system of multiple marking recommended by Wiseman (1949, 1956).

When used as criteria of creative thinking, imaginative stories would be more reliable if similar multiple marking methods were adopted, but with the labour involved and the experience needed, multiple marking is not feasible for a basic divergent thinking test. Scorer reliabilities in the region of 0.75 would consequently indicate too large a source of error, arising from scoring alone, to justify their use as psychological tests. If this degree of scoring error is found in the procedures established for the present study, it will limit seriously the likelihood of finding relationships between the test results and other measures.

This possibility is now to be investigated.

7.2 SCORER RELIABILITY IN THE PRESENT STUDY

Details of the scoring procedures used in this study have already been given in the last chapter. They are an attempt to provide a system of scoring which would validly reflect the qualities of divergent thinking being investigated and it has been argued that they should be more valid than some of those adopted by other investigators. The first consideration however is whether the procedures can maintain the high level of scorer reliability reported for other scoring procedures, particularly those of Torrance.

The investigation of scorer reliability was conducted within a stratified random sample consisting of the scripts of 42 pupils chosen from the total population of 11-year-olds (main sample).

7.21 Selection of the Sample

A sample of 42 pupils had already been drawn from the total population to form the frequency distributions of uncommon responses for the Uses and Consequences Tests, and, at first, this was also considered as a possible sample for investigating scorer reliability. The marking procedure for originality involves two distinct stages however, the drawing up of the frequency distribution and the scoring of the scripts. Scorer errors and differences of opinion can occur at both stages, though the second stage is less likely to cause difficulty if the scripts have already been scrutinised in order to draw up the frequency distribution. When only a sample is used to draw up the distribution however, new scripts that have not been used for this purpose can be an additional source of disagreement between scorers. If, therefore, the sample chosen to assess interscorer reliability was identical with the distribution sample, it would reduce the possible sources of error and exaggerate the subsequent reliability rating. It was decided, therefore, that a new sample would have to be constructed and that this sample should contain only the same proportion of scripts contributing to the distribution as would the complete population. With this condition added to the 'stratified' sampling requirement that each school class and sex should be represented by the same proportion in the sample as in the whole population, the pupils were randomly chosen using tables of random numbers.

7.22 Procedure

The scripts of each pupil in the sample for each of the three tests Uses, Consequences and Circles were marked by two independent scorers, the writer (scorer 1) and an university student (scorer 2). The whole of the scoring procedure, including the drawing up of the frequency distribution on a separate sample, was carried out independently by the two scorers. The procedures adopted were identical to those used in marking the whole population of scripts, except for the use of a sample of scripts to draw up the frequency distribution for Circles. Rather than ask the second scorer to draw up a frequency distribution for the whole population, the writer carried out the same procedures as the second scorer by marking the sample of scripts on the basis of a new frequency distribution constructed from the distribution sample. If anything the use of a sample of scripts for constructing the frequency distribution is likely to be slightly less reliable than using the whole sample so that the procedure adopted will not give an overestimate of the scoring reliability for the Circles Test.

Before embarking on the scoring the second marker was prepared for the task in discussion with the writer. The background to divergent thinking tests was discussed, the procedures explained, and a set of scored scripts from a previous study were analysed with their accompanying frequency distributions. Both scorers had access to the writer's list of flexibility categories for use with the Circles Test, as shown in the Appendix, but the principle explained earlier, of awarding new categories to responses which could not readily be related to the suggested list was adhered to.

7.23 Results and Discussion

Means, standard deviations and coefficients of interscorer reliability for the two scorers are given in Table 6.

The high coefficients of interscorer reliability more than confirm Anastasi's conclusion that with sufficient preparation such tests can probably be scored with satisfactory consistency, and all are greater than the arbitrary level of 0.90 suggested by Torrance.

The lowest coefficient, 0.91, is for the scoring of originality on the Uses Test and though less than the 0.98 values reported for the same test by both Wodtke (1963) and Yamamoto (1962), is very similar to the coefficient of 0.92 also

(a)

(n = 42)

		Scorer 1		Scorer 2		Coefficients of Interscorer Reliability
		Mean	s.d.	Mean	s.d.	
Circles Test	Fluency	19.12	6.74	19.00	6.63	0.993
	Flexibility	13.45	4.49	13.74	5.00	0.962
	Originality	18.62	14.80	17.43	13.93	0.957
Uses Test	Fluency	15.10	8.91	14.50	8.96	0.988
	Originality	14.24	10.92	15.48	10.8	0.910
Consequences Test	Fluency	10.93	5.16	10.90	5.09	0.991
	Originality	8.63	7.01	9.07	7.10	0.961

(b) Summary of Interscorer Reliability Coefficients

(n = 42)

	Fluency	Flexibility	Originality
Circles Test	0.99	0.96	0.96
Uses Test	0.99	-	0.91
Consequences Test	0.99	-	0.96

Table 6 Means, Standard Deviations and Coefficients of Interscorer Reliability between two Independent Scorers for three Tests of Divergent Thinking

reported by Yamamoto (1962). It is higher than many of the scorer reliabilities for originality reported by Torrance (1974), though it supports his observation that the scoring for originality usually causes the most problems. In Torrance's experiments however the poorest of his originality reliabilities were in scoring the figural tests. By contrast the Circles Test in the present study had a very high level of scorer reliability on all three of the abilities, fluency, flexibility and originality

As expected the scorers showed greatest agreement in scoring for fluency, reliability coefficients of 0.99 being achieved for each test. Even for fluency however the scoring is not as objective as some previous researchers have suggested by reporting coefficients of 1.00 (Yamamoto, 1962). This is also indicated by some slight differences in mean scores which reflect a tendency of the writer to be more generous than the second scorer in awarding credit to doubtful responses. This tendency is discussed later.

Scoring for flexibility in the present study was confined to the Circles Test, but the coefficient of 0.96 is greater than the 0.91 reported for the Circles Test by Yamamoto, and identical to the 0.96 reported for the same test by Wodtke. It lends weight to the arguments of the writer that a method of scoring based on previously recognised categories but also encouraging the award of separate credit to doubtful responses, can be scored consistently. In particular the present scoring method stresses that the scorer should not feel constrained to place 99% of the responses in the given categories as Torrance recommends. In spite of the differences between the two methods however, they yield very similar scorer reliabilities, Torrance, in the experiments reviewed earlier, reporting coefficients between 0.94 and 0.98.

This suggests that the error involved in the writer's greater reliance on the scorer to judge when to form new categories, is no greater than that involved in placing responses into Torrance's exhaustive list of categories; the latter, as illustrated earlier, often being too wide to be precise. The writer's claim to greater validity is not, therefore, hampered by a reduction in scorer reliability.

There is also no appreciable difference between the scorers in terms of the mean scores for each test, on any of the abilities tested. The largest difference is between the originality scores for the Uses Test, but although this is another indication of the problems of scoring for originality, the difference is far from statistically significant. The 't' value of the difference is 0.52, corresponding to a probability that differences as large as or larger than that recorded would occur by chance over 60% of the time.

Although when averaged over the whole sample there are no significant scorer differences, a scrutiny of the raw scores of individual pupils (given in the Appendix) can still reveal some quite wide variations between the scorers. A number of reasons become apparent after a re-examination of the scripts.

One or two random mistakes appeared in the totalling of marks and in the allocation of originality weights, but the major discrepancies between the scorers were more systematic and deserve further comment. They arose from two sources, firstly a slightly different interpretation of the principle suggested by the writer for deciding when to award marks for originality to responses that are nevertheless related in some way; and secondly a different degree of tolerance and appreciation of responses that at first sight might appear silly or irrelevant.

In general the principles of the scoring procedures suggested by the writer were well grasped by the second scorer, though the latter had a tendency to award credit for originality to individual responses when the writer tended to group certain responses together. For example, in allocating originality marks for Uses of a newspaper, the writer established three categories of response within the general area of use 'for lining things'. These were 'lining for a bin, bucket etc', 'lining for drawers, cupboards etc', and 'lining the floor of a cage as slip tray to help cleaning'. In addition there was a category of 'bedding to keep dog, cat etc. warm'. The second scorer gave more individual credit for responses such as lining for 'bin', 'bag', 'box', 'drawer', 'bucket', 'dog basket', and 'animal cage', as well as 'lining in general'. In the Circles Test responses showing the head of a 'pin', 'nail' or 'drawing pin' were grouped together by the writer but 'drawing pin' was given separate credit by the second scorer.

It is clear from these examples that no hard and fast rule is possible, though in subsequent discussion between the scorers closer agreement could usually be reached, both in the scoring of individual responses and on the principles underlying the scoring procedures. The latter is one area that could profitably be included in the "special attention" which Torrance (1974) notes "might be given to training in scoring for originality".

The second area of disagreement is a less technical one but reveals an important difference between the scorers, which, had it been greater, could have called into question the validity of the measures. It showed up in the extent to which the scorers gave credit for responses that at first sight appeared to be somewhat implausible or even facetious, and implied some variation in their basic views of the principles underlying the tests. In particular it showed in a different appreciation of the principles for scoring relevant responses discussed earlier. Although these principles, including the need for scores to be in sympathy with the ethos of testing divergent thinking, were discussed with the second scorer some difference of interpretation was revealed by a scrutiny of the marked scripts.

Superficially it might be suggested that the writer simply adopted a 'lenient' approach and the second scorer a more 'demanding' one, but the issue is more closely related to the nature of responses to divergent thinking tests than this implies. As argued in discussing the scoring procedures, it is essential to consider all responses seriously no matter how implausible they appear initially.

Some bias between the scorers appeared when their evaluation of responses was discussed in this light. For example, responses such as the following for 'Uses of a newspaper', involved some different interpretations: to make socks, for bandages, for milking a cow, draw a line, for patterns, hide dead bodies, a funnel, cigarette holder, and grow mustard and cress. In fact neither scorer could fathom the relevance of 'milking a cow', but the writer was more prone to give credit for the other responses, seeing some practical relevance in each of them. 'To make socks' for example was taken to refer to the type of practical use of a newspaper to keep feet warm inside Wellington boots, though the second scorer was somewhat sceptical of this interpretation. The difference of opinion illustrates the sort of dilemma that faces the scorer in having to make a judgement on the basis of such a small amount of information. As Wallach and Kogan (1966) point out, one of the advantages of an oral format is that it allows the scorer to ask pupils to elaborate on ambiguous or doubtful ideas, but this is not easy when the tests are set in a written form.

When evaluating written responses it is also more tempting for a scorer to interpret an obscure response as a facetious one if it is mis-spelt or badly written, than if presented as part of a lucid and neatly written statement. To use a newspaper to 'mack a skirt' is, for example, a less attractive response than one which states 'for making imitation grass skirts'. Although the responses are essentially alike it could be argued that the extra detail conveyed in the second deserves extra credit for originality. The better presentation inevitably creates a favourable impression on the scorer and may well influence his judgement of the quality of the basic idea. It has to be recognised that in a written form, divergent thinking tests are likely to measure a certain amount of general verbal ability.

Regan (1976) has also found that divergent thinking ability, when measured by verbal tests, is partly a reflection of the subjects' social class, though this is not so marked in the case of figural tests. Whether or not this is independent of the effect of intelligence was not determined in

Regan's study, though there was also a strong relationship between the divergent thinking tests and the verbal measure of I.Q. The latter cannot be completely avoided with written divergent tests, but in order to help emphasise their divergent nature it is essential for scorers to appreciate the value of ideas implicit in responses which are linguistically vague, inaccurate, or unrelated to their own social environment.

As noted earlier, Torrance (1974) goes as far as to caution scorers to score "any minimal response which can be reasonably interpreted or identified", and although, on the other hand, the writer would warn scorers of the danger of using their own imagination to interpret a response, the evidence of the present study reinforces the view that scorers should think twice before dismissing responses that are difficult to interpret.

Many responses become clear if given some extra attention, particularly if the scorer has had some experience of the problems, and common patterns, of children's handwriting and mis-spelling. For example, one girl, as 'Consequences of no need to eat' gave what appeared to be the following responses: "no more feed complates, no more farm or farmers, no more growing grops, no lauch Breaks needed, no need for eliets, animals would be wondering over the place, no more fannin, noone being posisioned". With appropriate insertions of the words, crops, lunch, diets, wandering, famine and poisoned, the responses take on a much more intelligible and competent flavour, though 'feed complates' is still a mystery!

The writer also found it reasonable to credit responses to 'Uses for a newspaper' such as 'lould spearks', 'to us as cortins', 'to put shoes on when you ploish them', and 'to put potoes peals in'; though he could not interpret 'to make Joatare'. For the Circles Test somewhat unintelligible objects labelled 'spird', 'plat', 'tadle', 'spy roll' and 'vunel' could also be interpreted on second thoughts, as could the response of 'no prasits' as a consequence of 'having no hair'.

It is argued that this 'interpretation' of responses is a valid procedure largely because of the degree of obviousness that the 'correct' interpretation conveys. This should however be apparent to others as well as the scorer, lest the latter become too disposed to take liberties in interpreting others responses. In the opinion of the second scorer there was a tendency for the writer to err in this direction at times, though generally agreement could be reached between the scorers in subsequent discussion. This supports the views of a number of researchers, (Wallach and Kogan, 1966; Barker Lunn, 1970; Vernon, 1971) that when a large amount of material has to be scored, satisfactory consistency can be achieved by a number of different scorers if doubtful items are resolved by discussion.

7.3 CONCLUSION

Prior to this study the writer was somewhat sceptical of the generally high levels of scorer reliability reviewed earlier, and concerned at the limited number of first hand reports of scorer reliabilities. Vernon (1971) in his study involving the scoring of over 900 pupil-hours of divergent test material points out that he omitted "the common practice of having each test scored by two persons", partly because of costs in time and money, and also because of the quite high interscorer reliabilities that have been reported. It is likely that this combination of reasons influences many researchers to rely on earlier evidence, though with the variety of scoring procedures and the inevitable degree of subjectivity noted earlier, it is surprising that there have not been more investigations of scorer reliability. It is also surprising in view of Vernon's (1971) comments that "There are still serious problems in deciding when responses are different or unusual", and "It is difficult ... to avoid some subjective evaluations".

It appears however that some of the writer's doubts were unfounded and that in spite of the subjective judgements involved, Vernon was correct in concluding that in his view such judgements are reasonable provided they are applied consistently and openly. The present evidence indicates that although making subjective judgements may weigh heavily in the mind of scorers, it does not influence the final scores as much as might be expected. The results support Hudson's (1966) suggestion that "Where errors or inconsistencies occur, these are rarely on a scale to threaten the value of the work as a whole".

On the other hand the variations in the scores awarded to certain individuals in the present study suggests that if information on particular pupils is required their scripts should be assessed in consultation with a second scorer whenever difficulties of interpretation arise.

In general the overall agreement between the scorers is very encouraging, especially considering the inexperience of the second scorer. The results show clearly that, although the scoring procedures may not be as objective as suggested by Cropley (1967), they are more than sufficient to allow the next stage of this investigation to proceed on a well-founded and reliable basis.

CHAPTER EIGHT

TEST-RETEST RELIABILITY OF TESTS OF DIVERGENT THINKING

8.1 INTRODUCTION

It was suggested in Chapter 1 that internal consistency methods are likely to give inflated measures of the reliability of divergent tests, and although administratively more convenient they should be replaced by test-retest correlations. It was also noted that while internal consistency methods are almost invariably high, test-retest correlations have often been unsatisfactorily low. The latter are now to be investigated for the three divergent tests used in this study.

Alternate forms of each of the tests were administered to the main population of 11-year-olds, about one week (± 1 day) after the initial testing.

The following tests were used:

Initial testing

Circles Test

Uses Test (Newspaper)

Consequences Test (No hair)

Re-testing

Squares Test

Uses Test (Spoon)

Consequences Test (No need to eat)

The re-test analysis was carried out on a stratified random sample consisting of the scripts of 65 pupils chosen from the whole population (n = 165). In view of the time involved in marking divergent tests this sample was considered large enough to give weight to the generality of the results. To have tested only this group however would have been to introduce considerable variations from the first testing and would also have been administratively less convenient. The whole population consequently took part in the re-testing, identical procedures being adopted, as already described in Chapter 5.

This sample was an extension of that constructed for assessing interscorer reliability. The latter (n = 42) was enlarged by choosing further pupils using random numbers, within the same constraints as for the former same. Thus each school class and sex contributed the same proportion of pupils to the sample as it did to the whole population, and the sample also contained the same proportion (one quarter) of scripts from the distribution sample (for originality) as did the whole population. Of the four pupils eventually omitted from the analysis for the reasons given earlier (page 209) two were members of the sample and the final number was therefore reduced to 63 subjects. This reduction was allowed to stand, rather than interfere with the procedures already adopted for selecting and marking the sample by introducing, and marking in isolation, two additional subjects.

Scoring procedures were exactly the same as for the original sample. This meant that in the case of the Squares Test the frequency distribution to allocate marks for originality was drawn up for all 63 members of the sample and the same percentage of responses credited as for the Circles Test. In scoring the two verbal tests the initial scoring procedure entailed using a frequency distribution for allocating originality marks which was based on a sample of scripts from the whole sample. Exactly the same sample was used for the scoring of originality in the retest.

8.2 RESULTS AND DISCUSSION

The test-retest reliability coefficients are presented in Table 7.

	Fluency	Flexibility	Originality
Circles Test	0.82	0.70	0.76
Uses Test	0.53	-	0.50
Consequences Test	0.75	-	0.60

Table 7 Test-retest (alternate form) reliabilities (n = 63)
(Interval between testings = 1 week)

These reliability coefficients can be evaluated in three ways, in terms of their improvement over chance in measuring divergent abilities, by comparing their values with those of other tests of the same type found in previous studies, and in terms of the proportion of total score variance which can be attributed to variation in true scores. Each of these ways will be discussed for the obtained coefficients.

8.21 Level of Reliability

There is no absolute level at which a reliability coefficient becomes acceptable and any test having significant reliability coefficients can provide some improvement over pure chance speculation. In the above table for 61 degrees of freedom, the values of the correlation coefficient are all significant ($p < 0.01$ for $r > 0.33$). It is however the size of the coefficient rather than its significance which is likely to influence decisions about the use of divergent tests. The reliabilities of 0.82, 0.70 and 0.76 for Circles and 0.75 and 0.60 for Consequences would be acceptable for many psychological tests particularly if used as sub-tests in an overall battery formed of similar tests. Reliability coefficients for sub-tests of the W.I.S.C., even by the split-half method are as low as 0.71 (Wechsler, 1949). The coefficients of 0.53 and 0.50 for the Uses Test however suggest that the test is of limited value where individuals are concerned.

Even so, they are not unlike sub-test reliabilities for personality scales. Test-retest correlations, after one month, for the Extraversion and Neuroticism scales of the Junior Eysenck Personality Inventory (Eysenck, S., 1973) range from 0.53 to 0.86 for children of 10, 11 and 12 years of age, and similar reliabilities are obtained by Eysenck and Eysenck (1973) for the new 'P.Q.' Personality Questionnaire. For the latter longer term reliabilities fall a good deal lower, down to 0.34, especially for younger children, yet the authors conclude that the reliabilities are not so low as to make it inadvisable to test young children for group comparison and for experimental purposes. They suggest however that as a general policy "correlations between personality and achievement, say, require to be corrected for attenuation before the 'true' correlations can be properly estimated". The use of the correction for attenuation will be discussed later.

Thorndike and Hagan (1969) provide a useful illustration of the accuracy of prediction resulting from tests with varying reliabilities. Supposing that two individuals 'A' and 'B' appear at the 75th and 50th percentile respectively after one testing, the chance that A will still surpass B if they were retested can be calculated from the forecasting efficiency of the test. With a reliability of 0.5 there is only a 2:1 chance that the order will remain as before, and even a reliability of 0.8 gives only a 4:1 chance. For individual assessment reliability coefficients consequently need to be very substantial.

For group prediction however comparatively low reliabilities show a marked improvement over chance. If the scores obtained by 'A' and 'B' were averages of a group of 25 people rather than individual scores the chance of a reversal for a test with reliability 0.5 is only 1 in 20 and for a reliability of 0.70 it is only 1 in 1000.

From the reliabilities in Table 7 it can be suggested that divergent thinking tests can play a consistent role in making decisions about groups though their use in identifying individuals is more doubtful particularly if just one test such as 'Uses' is adopted. Reliability is very like to increase however if the individual test scores were summed to form a composite divergent score. Whether the factorial structure of the correlations between divergent tests and their relationship with other variables makes them suitable for use in a composite score will be investigated later.

8.22 Comparison with other Studies

In spite of the comparatively low reliability of the Uses Test, the reliabilities are, overall, quite encouraging when compared with some previous researches. It should be recognised however that the range of ability in the present study maximises the chance of attaining high reliabilities.

In Dewing's (1970) study, already reviewed, the children were considerably above average in I.Q., and a year older than the children in this study, but her use of the Circles and Uses Tests provides a direct comparison. The reliabilities obtained by Dewing for each of the school populations used (n = 31 to n = 56), are summarised in the following table (Table 8).

Circles Test	Fluency	range 0.55 to 0.85	mean 0.69
	Originality	range 0.30 to 0.73	mean 0.54
Uses Test	Fluency	range 0.24 to 0.73	mean 0.52
	Originality	range 0.06 to 0.67	mean 0.39

Table 8 Test-retest Reliabilities (6-week interval) (from Dewing, 1970)

Considering that reliabilities are reduced when the range of ability is diminished and when the time between testings is lengthened, Dewing's mean reliabilities are quite consistent with those obtained in this study. The wide range of values, however, particularly for originality in the Uses Test, is somewhat disconcerting. More reliance could be placed on low and consistent reliabilities than highly fluctuating ones with only a very moderate mean value. Like those in this study Dewing's reliabilities are higher for the Circles Test, and for fluency as compared to originality. This is a common pattern in other studies. Yamamoto (1962) obtained reliabilities of 0.76, 0.63 and 0.79 for the Circles Test (fluency, flexibility and originality respectively), and 0.75, 0.60 and 0.64 for the same abilities on the Uses Test, though the population was considerably older, student teachers, and smaller, n = 22, than that in this study. Torrance (1974) reports similar values though usually for composite batteries of tests rather than for individual tests, and most frequently for high school or college students. Elsewhere however Torrance reports on a test-retest study with 118 5th grade children of similar age to those in this study and with an identical one-week interval (Torrance and Aliotti, 1969). Summing the scores for each ability, such as fluency, across the verbal and figural sub-tests the figural reliabilities ranged from 0.71 to 0.85, and the verbal reliabilities from 0.84 to 0.93. The former are very similar to those reported here though the latter even allowing for the effect of summing across tests are considerably higher. The trend for high figural than verbal reliabilities is also reversed.

In a smaller study with a similar time interval and 5th grade children, Hagender (1967) obtained the following reliabilities:

Verbal fluency (0.87)	Figural fluency (0.50)
Verbal flexibility (0.84)	Figural flexibility (0.63)
Verbal originality (0.79)	Figural originality (0.60)

The magnitude of the correlations is not unlike those obtained here, but once again there is a reversal between verbal and figural values, the verbal tests having similar reliabilities to the writer's figural test and vice versa. Each coefficient however is once again for a composite score across all the tests in Torrance's battery rather than for a single test.

Much lower reliabilities were obtained by Wodtke (1963) both for individual tests and for his combined battery based on Torrance's tests though the interval is one of 6 months. Wodtke's values, based on groups of between 100 and 150 children in both the 4th and 5th grades, range, as noted in Chapter 1, from 0.03 to 0.62 for the non-verbal sub-tests and from 0.21 to 0.74 for the verbal tests. Reliability coefficients obtained by Wodtke for the tests investigated here are given below in Table 9.

		(Grade 4 children) (n = 105 to 120)	(Grade 5 children) (n = 132 to 147)
Circles Test	Fluency	0.60	0.61
	Flexibility	0.32	0.37
	Originality	0.60	0.55
Uses Test	Fluency	0.28	0.38
	Originality	0.25	0.38
Consequences Test	Fluency	0.42	0.68
	(Flexibility)	(0.39)	(0.32)

Table 9 Test-retest Reliabilities - over 6 months

(from Wodtke, 1963)

Although the overall level is considerably lower, the pattern of Wodtke's results for these three tests is similar to the writer's. The Uses Test is the least reliable and the Circles Test, overall, the highest.

Wodtke considered that the higher reliabilities for the Circles Test may have resulted from the fact that identical forms of the test were used, there being no alternate form published by Torrance at that time. Even with the alternate form 'Squares' used in this study, the Circles Test is clearly the most reliable of the three tests used, and alternative explanations to that of Wodtke can be offered.

The most obvious explanation is that the variation is caused by different abilities being demanded by the tests, though the question of interest and motivation also needs to be considered. Children obviously enjoyed doing the Circles Test (and its alternative form) and frequently asked the writer to allow them to continue after the fifteen minutes allowed.

This applied right across the ability range, and the continued enthusiasm for the alternative form indicates that motivation can be sustained when retesting with this test. The writer feared in fact that the children's enjoyment of the test might mean that the responses could be so easily made that the test would fail to differentiate between individuals. The high correlation between 'Circles' and 'Squares' however suggests the opposite, some individuals consistently performing at a more fluent, flexible and original level than others.

The enthusiasm for the Consequences Test was less marked, though the procedure of giving each test separately and inserting it as another 'enjoyable activity' into the normal lessons, given by the class teacher, meant that the children appeared to react positively to each of the tests.

In dealing with 'implications' however the Consequences Test might demand more of the children's reasoning ability than the other two tests, and considering the range of ability in this study its consistency might owe something to its relationship with intelligence.

The Uses Test appeared to be somewhere between the other two tests in its level of popularity and this also appears to be true of its level of difficulty. As noted earlier it is the most common of all divergent tests, and in view of this its reliability is disappointing. With similar evidence from other studies, this may have to be regarded as realistic in the present stage of development of divergent tests.

The reliabilities for the Circles and Consequences Tests however are not unlike those reported for established tests. Though retest reliabilities for I.Q. tests are generally in the 0.80's and 0.90's (Anastasi, 1968) retest correlations for the sub-scales are often lower; Derner, Aborn and Canter (1950) for example reported reliabilities for the Wechster-Bellavue sub-tests as low as 0.64. For personality scales the reliabilities are markedly lower. For the extraversion and neuroticism dimensions of the Junior Eysenck Personality Inventory test-retest reliabilities calculated for separate age groups range from 0.51 to 0.91, the overall reliabilities for E and N being 0.6o and 0.75 for boys and 0.67 and 0.74 for girls respectively (Eysenck, 1973). In comparison, the reliabilities of 0.82, 0.70 and 0.76 for the Circles Test and 0.75 and 0.60 for the Consequences Test are quite encouraging, and though not particularly high in absolute terms, suggest that the tests deserve investigation in relation to other variables. The statistical limitations on any such relationship will be looked at later in this chapter.

8.23 <u>Levels of Performance from Test to Retest</u>

Although alternate forms are not identical to the initial tests, a comparison of level of performance from one testing to another can still indicate consistency of performance, and may cast further light on any difference in overall motivation from one session to another. Means and standard deviations of each test and its alternative form are given in the following table (Table 10).

There appears to be little indication of any reduction in motivation for the retesting. For both the Uses and Consequences Tests mean scores for fluency are increased and for the Circles Test they are almost identical on the two occasions. This information, however, should be interpreted in conjunction with the significant correlations between test and

	Initial Testing		Retesting		't' value of difference of means	Variance ratio 'F'
	mean	s.d.	mean	s.d.		
Circles						
Fluency	19.78	7.33	19.76	8.12	NS	NS
Flexibility	14.65	4.76	13.75	4.42	NS	NS
Originality	12.76	9.91	17.00	12.73	2.09**	1.65+
Uses						
Fluency	11.17	4.28	11.65	5.26	NS	1.51
Originality	9.49	6.59	9.81	7.13	NS	NS
Consequences						
Fluency	9.11	3.89	9.62	6.08	NS	2.44++
Originality	7.68	6.08	10.02	8.21	1.81*	1.82+

** values of $t > 2.00$ are significant $p < 0.05$ (2-tail test, 61 d.f.)
* values of $t > 1.67$ are significant $p < 0.10$ (2-tail test, 61 d.f.)

++ values of $F > 1.67$ are significant $p < 0.05$ (2-tail test, 61 d.f.)
+ values of $F > 1.53$ are significant $p < 0.10$ (2-tail test, 61 d.f.)

Table 10 Means and Standard Deviations for Divergent Tests (alternate forms) on two occasions one week apart retest performance and the amount of dispersion of the scores indicated by the standard deviation.

The standard deviation for fluency is greater for each of the alternate forms of the tests, very significantly so for the Consequences Test and nearing significance for the Uses Test. Children who are most fluent at the Consequences Test on the first occasion consequently tend to do even better at the second attempt while the reverse is true for those who do not score well in the initial testing. To some extent this is also true for the Uses Test. This finding is consistent with the writer's suggestion that the Consequences Test makes more demands on children's thinking than the other tests and is likely to receive less effort from some of the pupils on retesting. Whether or not this implies a greater degree of correlation with intelligence will be investigated later when the factor structure of the tests will be looked at in relation to other variables.

In the case of the Circles Test the standard deviation is not significantly different and for flexibility it is slightly reduced, as in the mean score. The originality score however is much increased together with the degree of variation. This is also true for the Consequences Test. In this case the conclusion made for the fluency performance on the Consequences Test is not applicable as the significant increase in variation is accompanied by an increase in mean score. This pattern, together with the correlation between first and second testings suggests that divergent thinking behaves like most abilities in showing an overall increase after practice, but also a preservation of relative ranking. This is a typical finding when I.Q.'s are retested over a short interval, as illustrated by the following analysis of data (Table 11) taken from Richards (1970).

	Group A (n = 102)		Group B (n = 71)		Group C (n = 91)	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
1st occasion	101.46	12.77	101.62	12.61	104.14	11.67
2nd occasion	105.34	14.32	103.97	14.36	107.46	12.62
't' value of difference	2.04**		1.04		1.85*	
'F' ratio of variances		1.26		1.30		1.17

** Significant at 5% (2-tail test)
* Significant at 10% (2-tail test)

Table 11 Test-retest performance for alternate forms of the Moray House U.R.Q. test (interval two months)

On retesting there is a significant increase in I.Q. in two out of three of the groups and, though not statistically significant, the standard deviation is also enlarged in all three cases. The pattern of originality scores for the divergent tests is very like this though it is slightly complicated by the significant increase in variance. While some increase in variation to accompany the increase in mean score is to be expected it may also be due in part to variation in motivation, particularly in the case of the fluency score for the Consequences Test.

In the case of convergent tests retesting can be complicated by memory or practice effects, though with the pressures on pupils to face up to a 'test', motivation is generally assumed to remain reasonably uniform. The open-ended nature of divergent tests however relies a great deal on the individual's motivation, and there are some suggestions in the present evidence that the initial testing experience may have differentially affected the children's approach to the second testing. This possibility deserves further study. Some relevant evidence may arise from the final part of this study when divergent thinking performance will be looked at in relation to other variables including measures of attitude and personality.

Apart from the slight evidence that divergent ability depends on affective as well as cognitive factors the overall pattern of reliability coefficients, means and standard deviations are not unlike those that would be expected from most cognitive variables.

8.24 Reliability Coefficients and Distribution of Variance

The level of reliability however varies between tests and for each test places a limit on any possible relationship with other variables or validity criteria. This aspect will be clarified by looking at the amount of error variance and 'true' variance in the test scores, and will have to be borne in mind when appraising correlations obtained with divergent measures later in this study.

As outlined in Chapter 1, reliability coefficients can be interpreted directly in terms of the percentage of variance attributable to true scores and that due to error; the sources of the latter depending on the method used to determine reliability.

Combining the results from the assessment of alternate-form reliability and scorer reliability, an estimation of error variance in divergent thinking scores can be made to cover errors over time, content and scoring. This is illustrated in the following diagram for the best and worst results obtained.

(i) Circles Test $r_{tt} = 0.82$. Scorer reliability $r_{s_1 s_2} = 0.99$

True variance = 81%	Error variance = 19%	
81%	Fluctuations over time and Content sampling 18%	Scoring variation 1%

(Not to scale)

(ii) Uses Test $r_{tt} = 0.50$. Scorer reliability $r_{s_1 s_2} = 0.91$

True variance = 41%	Error variance = 59%	
41%	Fluctuations over time and Content sampling 50%	Scoring variation 9%

(Not to scale)

Fig.7. Percentage distribution of variance for the Circles Test (fluency) and the Uses Test (originality)

The greatest amount of error variance in the present tests is therefore 59% and the least 19%. Unfortunately an inconsistent test cannot be a good predictor as the error portion of the test score will not correlate with any criterion. Consequently, the greater the error variance the lower any validity coefficient is likely to be. Theoretically a test cannot correlate with any outside variable more highly than it would with a 'true' measure of itself. The theoretical correlation of a test with 'true' scores, free from chance errors, known as the test's index of reliability, is the square root of the reliability coefficient, i.e. equals $\sqrt{r_{tt}}$ (Cronbach, 1970). Correlations of the test with any other variable or criterion measure (r_{tc}), must therefore be less than or equal to $\sqrt{r_{tt}}$, that is, $r_{tc} \leq \sqrt{r_{tt}}$.

Taking a coefficient of reliability based on the amount of true variance illustrated above, the highest possible correlation of the Circles Test with any other measure is $\sqrt{0.80} = 0.89$, and for the Uses Test $\sqrt{0.42} = 0.65$. While chance association of errors could inflate these theoretical maxima it is extremely unlikely provided the initial reliability estimates are realistic.

Adopting the test-retest method with alternate-forms over an interval of time provides the most rigorous method for assessing reliability and with the addition of error variance due to scorer variation, it is possible that the above results may overestimate the error term. In scoring divergent tests some error in scoring is likely to appear even with the same scorer, and this is already included in the test-retest assessment. On the other hand the scoring procedures laid down for the present study resulted in a high level of agreement between scorers, while an inexperienced scorer can find some difficulty in applying Torrance's scoring guide reliably. (See for example Regan, 1976.)

Whatever modifications could be made to the reliability estimates calculated above, they nevertheless fall well within the range of reported test-retest reliabilities and contrast markedly with those reliability estimates arrived at from internal consistency methods which, as noted in Chapter 1, are typically in excess of 0.85 (Wallach and Kogan, 1966; Cropley and Maslany, 1969; Zegas, 1976). Wallach and Kogan's split-half reliability for the fluency score for 'Uses' for example is 0.93 suggesting a limit of $\sqrt{0.93} = 0.96$ for correlations with other variables.

In the latter case failure to find a relationship of any magnitude with say, intelligence can then be interpreted as compelling evidence for discriminant validity. On the other hand if this internal measure of reliability is spuriously high, as it seems to be, then as Cronbach (1970) points out, how correlations of divergent tests with other variables can be attributed to their low reliabilities.

It would be difficult to fully justify this suggestion for the Circles Test or even the Consequences Test, but it is a realistic assessment of the Uses Test.

While discriminant validity demonstrated for tests of divergent thinking may, therefore, be partly a statistical artifact, the same may be true for any lack of convergent or criterion related validity. The true variance associated with the Uses Test is not sufficient for a very high level of correlation to appear with any other variable and what does appear may therefore be given less attention than it really deserves. It is rather like a type I error in testing a null hypothesis, one may be overlooking an important relationship by demanding too high a level of significance.

Having established the reliability of the divergent tests used here, it will be possible to make an allowance for the known proportion of error variance, and to estimate the correlation with other variables which would be likely to arise if the tests were perfectly reliable. This correction for the reduction or attenuation of correlation coefficients due to unreliability in test scores will be applied later in this study in connection with long-term stability and predictive validity.

8.3 SUMMARY

The test-retest (alternate form) reliability coefficients reported in Table 7 have been looked at in three main ways. It has been suggested that the size of the coefficients, which varied from 0.82 for the Circles Test (fluency) to 0.50 for the Uses Test (originality) are sufficient to indicate that divergent thinking tests can play a consistent role in making decisions about groups of children, though their use in identifying individuals is more doubtful. Unfortunately the test most often used to differentiate between convergers and divergers, the Uses Test, has the lowest reliability and results using this test should be treated with some caution.

Although internal consistency estimates of reliability are frequently in the region of 0.90 the present test-retest coefficients are considerably lower, though in accord with those reported by other studies of test-retest reliability. The writer suggests that the latter method gives a more realistic estimate of the reliability of divergent tests.

In general the pattern of correlations, means and standard deviations wasssimilar to that which would be expected of most cognitive variables, though an increase in variability on some measures suggests that test-retest reliability for divergent tests may be complicated by the introduction of motivational effects, not only in prefering certain tests, such as Circles, but in the way different children react to a retest situation.

Reliability has finally been looked at in terms of the percentage of variance attributable to 'true' scores and that due to errors. Correlations with other variables can only be done to the amount of non-error variance possessed by a test and this needs to be recognised when correlations of divergent thinking tests and other variables are presented in the remaining chapters.

CHAPTER NINE

LONG-TERM STABILITY OF DIVERGENT THINKING

9.1 BACKGROUND

In July 1969, in a previous study (Richards, 1970), several divergent thinking tests were given to 265 11-year-old children from three Primary Schools in the North of England.

Two of these schools were on the outskirts of a large city, and the third in the suburbs of a nearby town. The latter had placed special emphasis on mathematics teaching for over four years and was the 'experimental' school for the purpose of that study, while the former two schools, similar in most other respects were used for comparison.

Apart from the special attention given to mathematics in the one school, the children as a group formed a very representative sample of the whole population of the area. Excluding the children from the experimental school, the remaining sample is even more typical, and these 173 children form a natural group of reasonable size for a follow-up study. Even with the reduction in numbers to be expected over a period of five years, the final sample should be large enough to be amenable to statistical analysis and to give some weight to the generality of the results. Details of the age, I.Q. and social class of the children as initially found in their two schools are given in Table 12.

The problem in any educational study and in long term studies in particular is that results are influenced by an infinite number of variables having a differential effect on the individuals being studied. Most of these variables are

	School A	School B
Number of pupils	102	71
Mean age (s.d. in parenthesis) (at 1st July 1969)	11y 3.8m (3.3m)	11y 3.6m (3.6m)
Mean I.Q. (s.d. in parenthesis)	101.5 (12.8)	101.6 (12.6)
Percentage distribution of social class:		
Registrar-General's Classification I/II	12	17
III	65	64
IV/V	23	19

Table 12 Details of Initial Sample (1969) n = 173

beyond the control or even the knowledge of the researcher. Pupils' aspirations, attitudes, interests, relationships with parents and teachers, degree of support and encouragement, home background, and so on, all vary so enormously that one might be dubious about finding consistent patterns of personality or ability over any length of time.

The fact that some abilities such as intelligence do appear to be stable over long periods in spite of variations in people's experiences gives weight to their psychological, social and educational significance. The wider the range of some ability in a population however, the more likely it is to be detectable after some length of time, and if, as in the case of I.Q. measurements, they are themselves responsible for different treatment being given to high and low scorers, the stability is open to objections as being a self-fulfilling index rather than an underlying trait. The answer is likely to be a matter of degree, though it is obvious that in the case of I.Q., variations in pupils' experiences, particularly in different types of schools are likely to enhance rather than diminish its long-term stability. Any study of the reliability of I.Q. tests should therefore look closely at the nature of the sample used, especially if it constituted a collection of sub-groups selected on the basis of the variable being investigated.

With divergent thinking abilities the situation is somewhat different. At Primary School level it has been suggested that different types of teaching styles can be identified (Gardner, 1966; Bennett and Jordan, 1975) and that such differences in approach can affect children's divergent thinking abilities (Haddon and Lytton, 1968). In the Secondary School however Lytton and Cotton (1969) found it impossible to identify any homogeneous teaching patterns, and concluded that the variation in styles adopted by individual teachers result in there being no readily identifiable situational affect at the secondary stage.

Although over the whole range of intelligence there is inevitably some correlation between divergent thinking and I.Q. scores, especially in the verbal tests, any sample with a restricted I.Q. range is likely to have a considerable mix of divergent abilities. It is unlikely therefore that any long-term stability of divergent thinking abilities will be due to differential effects of school groups selected by I.Q. or academic ability.

In contrast to I.Q. therefore, it appears that divergent thinking abilities are unlikely to be systematically influenced by teaching methods or organisation in the secondary school, and it is feasible to investigate their stability over the whole population suggested. Any subsequent evidence for reliability over a long period of time will be all the more significant.

Following the raising of the school leaving age, none of the children tested in 1969 could leave school before the Easter of the school year in which they became sixteen, i.e. Easter 1974. It was decided therefore to try to make arrangements to carry out the follow-up testing during the school year 1973-74. Having obtained the permission and guidance of the Local Education Authority most of the children were traced to one Grammar School and one Bi-lateral School. After an interval of nearly five years it was satisfying to locate 150 of the original 173 pupils within these two schools.

The exercise was not without its traumas however as the school containing the largest proportion of the children was the last school visited out of four possible schools. In fact, after two schools had been visited, only one pupil had been traced. Including this pupil, who was allowed to join one of the testing sessions in another school, a total of 151 children out of the original 173 were eventually traced. Of these 139 were present and able to attend for the subsequent testing. Details of this final sample are shown in Table 13.

Number	139 (63 boys, 76 girls)
Mean Age (s.d.)	15y 9.8m (3.47m)
Mean I.Q. (s.d.)	101.51 (12.00)

Table 13 Details of Final Sample

It is important to consider whether the reduction of about 18% from the original sample is a random one across all ages and abilities or whether it introduces some systematic bias into the sample. In the only other comparable study reported in this country (Haddon and Lytton, 1971) a close scrutiny of the data reveals a bias in the sample which was not discussed by the writers. In an earlier study, Haddon and Lytton (1968) noted that in 1965, "211 children, 11-12 years old, were tested and the whole ability range in the schools was covered". In 1969 "as many as possible of the initial sample were traced to their secondary schools where they had completed three years and two terms" (Haddon and Lytton, 1971). Of the original 211 children, 151 were traced, who had then "attained, or nearly attained the statutory school leaving age". What the writers do not comment on is that out of the 25% who could not be traced a large proportion would no doubt have exercised their option and already left school. It is also likely that these pupils would have been the less academically able group and the resulting sample would not have covered the whole range of ability present in the 11-year-old sample. The mean I.Q. and age of the new sample was not given, though of the 148 children finally tested, 52 came from Grammar Schools and 96 from Secondary Modern Schools - over 35% from Grammar Schools.

Comparing the age and I.Q. of the initial and final samples in this study indicates that the reduction in numbers has not affected the overall balance of the sample, the mean I.Q.'s being almost identical and the standard deviation being only slightly reduced. The testing took place in early January 1974, four years and six months after the original testing and the mean age (less four years, six months) is, like I.Q., almost identical in the two samples. The findings confirm the results of the writer's general enquiries into the educational progress of children in the area over the five years in question. There was nothing to suggest that the children who could not be traced came from any particular group; there had, for example, been no marked change in housing conditions, such as large scale demolition; or in the employment situation, such as factory close-down, and the regulations governing school leaving ensured that none of the children concerned could officially leave school before the Easter following the testing. Some of the absentees may have considered themselves as unofficial leavers, but there were absentees from both Grammar and Secondary pupils, and apart from this possible explanation for the slight reduction in variance there was no marked effect on the final sample.

9.2 TESTS AND PROCEDURES

9.21 Divergent Tests

The three divergent tests investigated in this study, Uses, Consequences and Circles, were chosen, as described in Chapter 5, partly because of their widespread use in other investigations and partly because of their prior inclusion in earlier work of the writer. In this follow-up study they were administered in an identical format and with the same time limits as in 1969. Details of the tests and the scoring procedures have already been given and the tests themselves are reproduced in the Appendix. When discrepancies arose between the present recommendations for scoring (adopted with the new 11-year-old investigation) and that applied in 1969, the latter were adopted. This occurred mainly in the percentage of uncommon responses credited with originality marks and has already been detailed when scoring for originality was discussed in Chapter 6.

The Circles Test was scored for fluency, flexibility and originality and the two verbal tests for fluency and originality.

The distinction between figural and verbal scores has been maintained throughout, though the three figural and four verbal measures were standardised and added to give additional composite scores for figural and verbal measures of divergent thinking. This was carried out for the data from both the 11-year-olds and 15-year-olds. In the 1969 investigation the Uses Test was also scored for flexibility, but this score correlated 0.95 with the fluency score and was omitted from the retesting procedures.

9.22 Intelligence Tests

Two intelligence tests, Moray House Verbal Reasoning Test 81, and Test 82 (1968) had been given in 1969 as part of the County's 11-year-old transfer procedure, and the results had been made available to the writer. In view of the important part that intelligence might play in divergent test performance its effect will be investigated. To sum the two I.Q. scores to give an average I.Q. for each child however would increase its reliability and its predictive effects, and the intelligence score adopted here is therefore that of the first test (designated as I.Q.1 in the 1969 investigation).

9.23 Administration

The pupils were located in three main buildings, the grammar school, and two separate parts of the bi-lateral school. One testing session, involving 40 minutes of testing time took place in each school, the pupils being allowed out of their normal classes. In two cases the sessions took place in a working area of the school library, and in the third in the school hall.

All three sessions were administered by the writer, though in the case of that in the school hall the deputy headmaster, who had arranged the seating, was also in evidence. Adopting an encouraging manner the writer introduced the testing by reading informally the following 'introduction' which appeared on the test booklet:

> This is not an examination, it is part of a SURVEY to find out how good pupils are at thinking up new and interesting ideas.
>
> Some years ago a similar survey was carried out in junior schools and most of you were involved then, now we want to see how you do when you are older.
>
> There are no right or wrong answers so write down *as many ideas as you can think of*. Work quickly, each part will be timed.
>
> If you need more space continue your answers *opposite* on the back of the previous page.
>
> DO NOT TURN OVER UNTIL YOU ARE TOLD.

Each test was introduced separately as described in Chapter 5, and the complete testing booklet is reproduced in the Appendix.

Having located and gathered the pupils together it was tempting to widen the scope of the investigation. Even for the procedures adopted here however the writer had caused considerable upheaval in the schools, the pupils being spread over a range of classes and subjects. With C.S.E. and 'O' levels less than a term away, and a number of visits already having taken place to set up the arrangements, it might have interfered with the cooperative way in which the present investigation took place, had the writer attempted any extension. It would not be impossible however to locate the pupils again at some future point in their careers.

9.3 RESULTS

9.31 Introduction

Complete data for each of the 139 pupils, including their sex (63 boys, 76 girls) and whether they were in 'grammar' or 'secondary' streams is available from the writer. It has been suggested that unlike considerations of the stability of intelligence, the effect on divergent thinking of types of school and internal organisation is not likely to be systematic and the results are pooled for this investigation. Whether to combine the data for boys and girls is a more open question.

Cronbach (1968) argues that until definite trends to the contrary are suggested it is more profitable to consider children's thinking abilities in a combined population. With the exception of Vernon (1972) previous researches have often remarked on the similarity between the sexes on divergent performance (e.g. Cropley and Maslay, 1969; Dewing, 1970; Bennett, 1973). Others have conventionally adopted combined populations (e.g. Haddon and Lytton, 1971; Hargreaves and Bolton, 1972). While the inclusion of personality variables would incline the writer to Vernon's view, only cognitive variables are included in this section. Measures taken to look at this question more closely will be returned to shortly.

Overall the statistical nature of the data is such that the total number of cases (n = 139) can be regarded as a fairly large sample, the scores for the same test constitute independent pairs, they are essentially continuous, and they result in unimodal distributions which are sufficiently symmetrical not to undermine the assumption of a linear relationship between the variables which underlies the use of the Pearson product-moment coefficient of correlation. Both zero-order Pearson product-moment coefficients and first-order partial coefficients will be calculated to express this relationship.

Means, standard deviations and intercorrelations between the variables are given in the Appendix and these will be extracted for discussion where appropriate.

To help clarify the question of sex differences a nominal dichotomous variable (boy 0, girl 1) was also included in the correlation analysis purely on an exploratory basis. At least with only two categories such a distribution lends itself to questions of linear correlation more appropriately than a three category variable, but no claims are being made for precise levels of the correlations obtained. Its use however revealed an interesting variation in boy/girl effects on the variables under discussion which was felt worthy of further investigation.

9.32 Sex Differences in Divergent Thinking - 1969 and 1974

The following table (Table 14) presents the pattern of correlations between sex and both divergent thinking variables and I.Q.

Circles I (1969)	Flu.	0.04	Circles II (1974)	Flu.	0.06
	Flex.	0.07		Flex.	0.11
	Orig.	-0.01		Orig.	-0.10
Uses I (1969)	Flu.	0.21	Uses II (1974)	Flu.	0.27
	Orig.	0.11		Orig.	0.18
Consequences I (1969)	Flu.	0.20	Consequences II (1974)	Flu.	0.29
	Orig.	0.11		Orig.	0.13
I.Q. (1969)		0.10			

<u>Table 14 Pattern of Tentative Correlations of Divergent Tests and I.Q. with Subjects' Sex (boy 0, girl 1)</u>

Values of $r > 0.174$ are significant at 5% for 125 degrees of freedom and give an indication of the likely significance of the above correlations. Only five out of the fifteen coefficients exceed this value and they form a clearly separate group from the other correlations which are well below this level. They are also distinguishable in psychological terms, deriving from the fluency scores for the verbal divergent tests, Uses and Consequences, in both 1969 and 1974, and the verbal originality score for Uses in 1974.

Had there been an equal correlation with the verbal intelligence quotient it would have been tempting to explain this effect in terms of verbal reasoning superiority of girls. The correlation with intelligence however is small and non-significant, a fact that is borne out by the following mean I.Q.'s of the boys and girls separately.

(i) Boys n = 63 Mean I.Q. = 100.25 (s.d. = 11.3)

(ii) Girls n = 76 Mean I.Q. = 102.55 (s.d. = 12.5)

(No significant difference, t = 1.14)

While the slight I.Q. superiority of the girls is likely to be part of the explanation for their superior performance on some of the verbal divergent thinking measures, the latter may also reflect a more persistent effort by girls when written responses are required.

The effect of intelligence will be controlled for later, but in the meantime the consistent differences in performance between the sexes on the verbal fluency scores from 1969 to 1974 could result in any long-term stability being interpreted as a consistent boy/girl group difference in divergent thinking rather than as a matter of individual ability. This will depend on the level of stability reported, but even though the correlations with sex are not great, they raise sufficient doubts to suggest that long-term stability should be investigated within each sex as well as in the whole population.

9.33 Long-term Test-retest Reliabilities 1969-1974

9.331 Pearson product-moment correlations were therefore calculated between the results obtained in 1969 and 1974, for boys and girls separately and for the whole sample. Coefficients of stability for the seven divergent thinking measures, for combined scores for each test, and for an overall verbal score, are reported in Table 15. Combined scores were arrived at by standardising and adding the individual scores for each pupil.

The table as a whole provides quite an impressive set of reliabilities after 4½ years. The highest coefficient, 0.57, is for the combined score of the two verbal tests, though all the coefficients, including those for the separate sexes, are greater or equal to 0.35.

Although there are some slight variations between the

	Whole population (n = 139)	Boys (n = 63)	Girls (n = 76)
Circles. fluency	0.47	0.36	0.55
flexibility	0.44	0.38	0.47
originality	0.41	0.48	0.37
Combined Circles	0.49	0.46	0.51
Uses. fluency	0.46	0.35	0.47
originality	0.46	0.35	0.50
Combined Uses	0.50	0.39	0.53
Consequences. fluency	0.44	0.48	0.35
originality	0.40	0.41	0.37
Combined Consequences	0.47	0.48	0.42
Combined Verbal	0.57	0.55	0.55
	all significant for $r > 0.30$ $p < 0.001$	all significant for $r > 0.32$ $p < 0.01$ for $r > 0.41$ $p < 0.001$	all significant for $r > 0.30$ $p < 0.01$ for $r > 0.38$ $p < 0.001$

Table 15 Coefficients of Stability for Divergent Tests, 1969-1974

sexes they are not consistently in either direction. While the reliabilities for the Uses Test (fluency), for example, are 0.35 (boys) and 0.47 (girls), the corresponding values for the Consequences Test are 0.48 (boys) and 0.35 (girls). The level of long-term stability is in fact very similar for both boys and girls and the overall coefficients cannot therefore be interpreted simply as a measure of group differences between the sexes.

The coefficients for the whole population, while similar in magnitude to those from the single sex groups, are more significant due to the increase in the number of degrees of freedom. The single sex coefficients are all significant $p < 0.01$, but those from the combined population are significant at the 0.001 level. Adopting Cronbach's (1968) recommendation to combine data from the sexes unless there are substantial grounds for not doing so, the stability coefficients will therefore be looked at more closely for the population as a whole.

9.332 Comparison with Previous Researches

Direct comparisons are possible with the researches of Haddon and Lytton (1971) and Cropley and Clapson (1971), noted earlier, both studies having looked at the stability of divergent tests over a similar 4 to 5 year period between the ages of 11 and 17.

Cropley and Clapson reported reliabilities for the Circles and Consequences Tests, for a sample of 110 seventeen-year-olds, five years after the initial testing. Cropley's scoring procedure has been discussed earlier and although he used only an originality score in the above study it is very similar to that adopted here. He also compared both boys and girls separately and together. While Cropley applied some correction for restriction of range (the mean I.Q. of the sample was 119) the present sample is representative of a complete ability range. The parallel results between this study and that of Cropley are compared below in Table 16.

	Circles			Consequences		
	Boys	Girls	All	Boys	Girls	All
Cropley and Clapson (1971)	0.48 (n=57)	0.40 (n=53)	0.44 (n=110)	0.58 (n=57)	0.33 (n=53)	0.45 (n=110)
The present study	0.48	0.37	0.41	0.41	0.37	0.40
	(n=63)	(n=76)	(n=139)	(n=63)	(n=76)	(n=139)

Table 16 A Comparison of Long-term Reliabilities for the Circles and Consequences Tests scored for Originality

There is a remarkable degree of similarity between these results and this adds considerably to the confidence that can be placed in them. The findings of isolated experiments are naturally treated with a good deal of caution and it is encouraging to find independent investigations yielding similar results.

Although Cropley's coefficients are slightly higher than those in the present study the numbers in his experiment are smaller and the net result is that the results of the present study are slightly more significant.

All the coefficients are significant at the 1% level except for that of the Consequences Test for girls in Cropley's study which is significant at 5%.

A comparison with Haddon and Lytton's (1971) investigation is also possible (Table 17) though their composite figural and verbal scores are arrived at from a variety of scoring procedures, discussed earlier in Chapter 6, and from two figural tests and three verbal tests. One of the figural tests, however, was the Circles Test, and one of the verbal tests 'Uses', and this improves the comparability of the results. Haddon and Lytton's final population consisted of 148 children first tested at 11 years of age and again four years later.

	Figural	Verbal
(n = 148) Haddon and Lytton (1971)	0.41	0.57
(n = 139) The present study	0.49	0.57

Table 17 A Comparison of Long-term Reliabilities for Figural and Verbal Tests of Divergent Thinking

Once again Table 17 shows a very high level of agreement and gives weight to the generality of the present results. There seems little doubt that divergent thinking has a very consistent, if moderate, degree of long-term stability.

The values of the coefficients are lower than those of Dalbec (1966), but her 3-year reliabilities of 0.59, 0.35, and 0.73 for the fluency, flexibility and originality scores of Torrance's tests, were obtained with a population of college students (n = 43), and considering her more mature population and smaller numbers the difference is not as marked as appears at first sight.

Although the present stability coefficients are not high, the property of stability, as pointed out in Chapter 1, should be associated with the underlying trait rather than the test itself, and the coefficients should not be evaluated in direct comparison with short-term reliabilities. For established tests of mental abilities the latter, as noted earlier, usually fall in the 0.80's and 0.90's (Anastasi, 1968), but these should not be used as a yard-stick to judge long-term data.

As noted in Chapter 1 test-retest reliability usually decreases as the interval between testing gets larger. Cronbach (1970) retesting 200 y-year-old children reported reliabilities for the Stanford-Binet scale of 0.91 after one week reducing to 0.74 over four years and 0.68 after eleven years. Also reporting on the Stanford-Binet scale, Thorndike and Hagan (1969) present somewhat lower retest correlations. In comparison with I.Q. assessed at around 17 years of age, I.Q. assessed two years earlier correlated in the region 0.70 to 0.80, tests taken from three to six years earlier around 0.60 and seven or eight years earlier around 0.50.

It should also be recognised that, in contrast to the composite score given by standardised tests based on a large number of sub-scales or groups of items, the results of divergent thinking tests are more comparable with those of sub-tests of standardised scales, unless added to form an overall divergent score. Cropley (1964) reporting two-year reliabilities for the sub-tests of the Wechsler (W.I.S.C.) Scale gave reliabilities ranging from 0.12 to 0.50 and the present reliabilities over a 4½ year interval fall at the higher end of this range. They are also higher than several of those reported for the new 'P.Q.' Personality Questionnaire

(Eysenck and Eysenck, 1973) after a 1-year retest, though generally the latter are in the 0.55 to 0.65 region.

In assessing long-term reliabilities Cronbach (1970) points out that reliabilities are likely to be higher over a period of little developmental change in the subject, in comparison with a time interval over important formative years. The period from 11 to 16 years of age covered by the present study is clearly in the latter category and the degree of stability reported is thus all the more significant.

In considering the development of divergent thinking ability over the period in question it is also appropriate to consider whether over and above the relative ranking of the subjects, there is any change in the level of their performance.

9.34 Levels of Performance on Divergent Tests 1969-1974

It is suggested that like other mental abilities divergent thinking should increase as children develop between the ages of 11 and 15, and this is clearly shown by comparing the mean scores in Table 18.

Putting the above statement, suggesting a development in divergent thinking ability, into the form of a null hypothesis and applying a t-test for correlated pairs (2-tail), each of the differences in Table 18 shows a highly significant increase well beyond the 0.1% level in all cases.

The children concerned were therefore more fluent, flexible and original in their responses at 15 than they were at 11. One could expect an increase in the fluency score partly on the grounds of writing speed, but an increase in fluency

Test		11-year-olds		15-year-olds		Difference
		Mean	s.d.	Mean	s.d.	
Circles	Fluency	19.27	7.60	23.81	6.74	4.54**
	Flexibility	13.91	5.56	19.34	4.93	6.43**
	Originality	6.26	6.48	14.72	11.12	8.46**
Uses	Fluency	23.26	15.79	31.19	10.66	7.93**
	Originality	22.10	10.66	30.16	19.06	8.06**
Consequences	Fluency	9.78	3.45	18.74	6.80	8.96**
	Originality	7.14	3.98	9.97	5.12	2.83**

** Significant at the 0.1% level. t-test for correlated pairs (2-tail), n = 139.

Table 18 Means and Standard Deviations of Divergent Thinking Scores at 11 and 15 Years of Age does not necessarily imply an increase in flexibility or originality. Remembering that the latter is assessed relative to its own particular population it is appropriate to infer a real improvement in each ability.

The increase in standard deviations from 11 to 15 are consistent with the increases in mean scores, though it is noticeable that in the case of the Circles Test for fluency and flexibility, and the Uses Test for fluency, the standard deviation is reduced. Not only, therefore, is average level of performance raised but the group is more closely grouped around the mean. Taking a complete ability range it is likely that at eleven years of age some of the children are unduly handicapped by limited writing skills, but by the age of 15 this variation in psychomotor skill is considerably reduced.

Taking the increase in the level of divergent thinking together with the consistency in relative ranking as demonstrated by the stability coefficients, it is reasonable to conclude that divergent thinking abilities develop with motivation in a similar way to other mental abilities. Whatever their relationship to other variables, the results therefore suggest that divergent thinking tests measure a trait which shows some definite stability, and reasonable development over time.

9.35 Correction for Attenuation

The level of stability is only moderate but it has been demonstrated in a number of other studies and suggests that divergent thinking could be a useful factor in predicting future behaviour. As the instruments used to assess the stability of divergent thinking are not perfect however, it is possible to consider further the maximum theoretical stability of the underlying construct that might be achieved were the tests more reliable.

In discussing reliability in the last chapter it was noted that unreliability in test scores inevitably reduces the correlation between variables but that a 'correction for attenuation' can be applied to give the maximum value of the correlation that could be expected if the tests were free from chance errors. It must be remembered however, that this is a theoretical not an empirical value and that it should not be used to make a test appear to be more reliable than it actually is. A bare statement of reliability therefore has to reflect the limitations of the test and the corrected values to be calculated here are not being suggested as an alternative to the empirical values of stability already established. They can however provide additional information about the construct underlying the test by making allowances for the shortcomings of the test itself. Correlations corrected for errors of measurement give a better notion of the intrinsic relationship between the variables (Garrett, 1966).

Thorndike and Hagan (1969), in an example of the value of the correction, assume that a test of arithmetic and one of reading correlate 0.56 before correction for attenuation and 0.70 afterwards, and they conclude that "in thinking of these two *functions*, it would be appropriate to think of the correlation as .70 rather than .56, though the *tests* correlate only .56" (original emphasis).

It is sometimes difficult, as Guilford (1973) observes, to decide on the value of the reliability coefficient to be used in correcting for attenuation and it is possible to under-correct or over-correct. Guilford suggests that an alternate-form coefficient, with no time between test and retest, probably the best to use for a conservative correction, but that whatever estimate is adopted the interpretation of the results will have to be made accordingly.

The 'net' reliability estimates arrived at in the last chapter were designed to include all the major sources of error and are likely, as noted then, to be slightly over-rigorous estimates. In correcting the long-term stability coefficients however, no allowance for scorer variation will be made as both sets of tests were marked by the same scorer and any errors from this source are already included in the test-retest coefficients.

The test-retest reliability of the tests for the 15-year-old population are not known, though as suggested earlier the evidence of other studies points to increasing reliability with age. This is based however on evidence from composite scores over a number of tests, and no specific reliabilities for the present tests are available for a comparable 15-year-old population. The known 11-year-old reliabilities will therefore be adopted for both initial and final tests even though there is acknowledged that this is likely to overestimate the correction but it will enable an approximate estimate of the theoretical stability to be made.

The corrected long-term stability coefficients $r_{t_{\infty}t'_{\infty}}$ are given in the following table (Table 19).

		Basic stability coefficients r_{tt^1}	Short-term test-retest reliability r_{tt}	Theoretical max^m value of stability $r_{tt^1_{\infty\infty}}$
Circles	Flu.	0.47	0.82	0.57
	Flex.	0.44	0.70	0.63
	Orig.	0.41	0.76	0.54
Uses	Flu.	0.46	0.53	0.87
	Orig.	0.46	0.50	0.92
Consequences	Flu.	0.44	0.75	0.59
	Orig.	0.40	0.60	0.67

Table 19 Estimates of Maximum Long-term Stability Coefficients for 'True' Divergent Thinking Measures

The theoretical stability values show a marked increase over the obtained values and give further support to the presence of an underlying construct of divergent thinking that remains stable in relative ranking over a period of years.

Even after correction however the theoretical maximum values for the Circles and Consequences tests indicate that there is no likelihood of perfect agreement between divergent measures over this time. The considerable increase in the stability coefficient for the Uses Test suggests that either the proportion of variance which is stable over a short period of time is relatively more stable than that of the other tests over longer periods, or that the short-term reliability is an underestimate.

In the short-term retesting with the Uses Test may be more prone to error than the other tests. The fact that in spite of its lower reliability the long-term stability of the Uses Test is similar to that of the other tests gives some weight to the writer's suggestion that its retest reliability may be influenced by short-term attitudinal variables in a way in which the Circles and Consequences Tests are not. On the other hand the low level of reliability reported by the writer is not uncommon, and there are therefore some grounds for believing that, over the long term, the divergent thinking ability underlying the Uses Test is more stable than that of the other tests.

Even though the actual values of the theoretical stability coefficients are likely to be overestimated due to the choice of a stringent level of reliability, their general magnitude indicates that it is reasonable to expect divergent thinking tests to be able to play a considerable role in the long-term prediction of children's abilities.

The evidence presented earlier in this study however leads one to consider whether the long-term reliability demonstrated for the divergent tests is due in whole, or in part to its overlap with general intelligence. To investigate this possibility we turn to the original long-term coefficients and the parallel correlations of the divergent thinking tests with intelligence.

9.36 Relationship with Intelligence

That there is some overlap between divergent thinking and I.Q. can be seen from the following table (Table 20), giving the long-term reliabilities of the three divergent tests and the combined verbal score, together with the correlations of each measure with I.Q. as assessed in 1969.

	Divergent Thinking Tests, 1974			
	Circles	Uses	Consequences	Combined Verbal
Test-retest correlations	0.49	0.50	0.47	0.57
Correlations with I.Q. (1969)	0.20	0.35	0.42	0.43

Table 20 Correlations of Divergent Thinking Abilities (1974) with Corresponding Tests (1969) and I.Q. (1969)

The correlations between intelligence and the divergent thinking tests are particularly substantial for the verbal tests, though even the correlation of 0.20 between the Circles Test (1974) and I.Q. is significant at the 5% level. In each case however the divergent score obtained in 1969 is a better predictor than I.Q., of the corresponding divergent score in 1974.

The important question however is whether the long-term stability is a 'true' relationship between the divergent scores or largely the effect of both scores being related to general intelligence. It is possible to deal with this question by controlling for the effects of intelligence via the technique of partial correlation.

As Guilford (1973) points out, this is the basic way of dealing with a three (or more) variable problem - the sort of problem which McNemar (1964) maintains is so badly handled by researchers who, in 'half-blind' designs, look at the relationship between divergent thinking and a further variable without considering the effect of the intervening variable of intelligence. While Guilford suggests that partial correlation techniques are most useful for partially out variables such as chronological age and intelligence, there is a danger of over-compensating by removing the effects of intelligence.

It would be difficult, for example, to interpret the relationship between tests such as mathematics and reading comprehension with the influence of general intelligence removed. Both verbal and mathematical understanding enter with heavy, but unknown, weight into most I.Q. tests and partial correlation between the two with I.Q. held constant could not be interpreted in any clear-cut and meaningful way (Garrett, 1966; Lewis, 1967).

To some extent the removal of I.Q. from any psychological test tends to remove more than it reasonably should, including a variety of test-taking abilities that it would be unreasonable to associate with I.Q. tests alone. The dice are consequently slightly loaded against new tests, though if they are to be worthwhile developing, they need to demonstrate their ability to predict relationships over and above those due to intelligence.

To assess the long-term stability of divergent thinking with the effect of intelligence removed, it is necessary to know the correlations between all three variables concerned. The correlations between I.Q. and the Circles, Uses, Consequences and Combined Verbal scores obtained in 1969 are 0.26, 0.52, 0.44 and 0.54 respectively. Taking these together with the correlations in Table 20 the effect of intelligence can therefore be removed from the long-term test-retest correlations. The result is to reduce the stability coefficients as shown in the following table (Table 21).

	Circles	Uses	Consequences	Combined Verbal
Zero order correlations	0.49	0.50	0.47	0.57
First order partials	0.47	0.40	0.35	0.45

Table 21 Long-term Stability Coefficients Before and After Being Controlled for the Effect of Intelligence

Even after this reduction there is still a significant relationship between divergent measures over the 4½-year period ($p < 0.001$ for $n - 3$ degrees of freedom) though in terms of common variance the reduction is more marked. It must be remembered that correlation is not a linear index and while the reduction in common variance between the Circles Tests in 1969 and that in 1974 is only down from 24% to 22%, $\left((0.49)^2 - (0.47)^2\right)$, that for the Consequences Test is reduced from 22% down to 12%.

Had this analysis taken place on the corrected reliabilities however, the reduction due to intelligence would have been proportionally lower, the correlations with the more reliable I.Q. test having less correction than those between divergent scores.

The effect of intelligence is so often suggested, but less often investigated, as a possible source of the relationship between divergent tests, that it was felt appropriate to examine its relationship to the long-term test-retest reliabilities. Having shown that there are no grounds for believing that the relationship is simply one of intelligence, the original analysis is able to stand more firmly as an indication of the stability of divergent tests.

9.37 Intercorrelations between Divergent Thinking Tests 1969 and 1974

In addition to the level and relative ranking of pupils' performance on individual tests from 1969 to 1974, an important aspect of the stability of divergent thinking is the relationship between the three tests which purport to measure the same construct, firstly in 1969 and then in 1974. These relationships are now examined. Intercorrelations between the tests are presented in Table 22(a) and 22(b) for the 1969 and 1974 data respectively. In 1969 the Uses Test was also scored for flexibility but this correlated 0.95 and 0.84 with the corresponding fluency and flexibility scores and was omitted from the 1974 analysis. For ease of comparison it is also omitted from the 1969 table.

The correlations between all the measures are high, and very similar patterns exist for both sets of data. This combination of both level and pattern is the most significant feature in the comparison and gives considerable additional weight to divergent thinking as a stable construct.

Some minor variations in the pattern are also worthy of note, the main difference being in the greater consistency in the 1974 results. Correlations between the two verbal tests (shown within the small squares in the table) ranged from 0.42 to 0.63 in 1969 and from 0.53 to 0.59 in 1974. The pattern is still the same however, the higher correlations existing between fluency scores and the lower ones between originality scores. Looking at the rectangular block of correlations between the verbal tests and the Circles Test, the 1969 correlations varied from 0.41 to 0.61 and the 1974 correlations from 0.46 to 0.55. The variation is most marked

(a) 1969

		1	2	3	4	5	6	7
Circles	Flu.1	-	86	64	61	51	48	42
	Flex.2		-	79	64	58	57	49
	Orig.3			-	47	41	43	43
Uses	Flu.4				-	81	63	49
	Orig.5					-	52	42
Consequences	Flu.6						-	79
	Orig.7							-

(b) 1974

		1	2	3	4	5	6	7
Circles	Flu.1		92	67	53	47	50	50
	Flex.2			65	51	46	52	47
	Orig.3				49	55	49	47
Uses	Flu.4					86	59	58
	Orig.5						53	53
Consequences	Flu.6							78
	Orig.7							

Correlations > 0.30 are significant $p < 0.001$

(Decimal points are omitted in the tables)

Table 22 Intercorrelations between Divergent Scores in 1969 and 1974 by the low correlation in 1969 between the originality score for Circles and the other measures. It suggests that there is some difference in the type of performance at 11 and 16 on the Circles Test and this is substantiated by looking at the correlations within the Circles Test. (Within-test correlations are enclosed by triangles.) Originality has similar correlations with fluency of 0.64 and 0.67 in 1969 and 1974 respectively but has a considerably higher correlation with flexibility in 1969 than in 1974. It appears that in 1969 the flexibility score played a more intermediary role between fluency and flexibility than in 1974 when it was closely related to fluency. It suggests that in 1974 the older children tended to change category more consistently rather than persevere with a number of responses in the same category. This in fact reflects the type of development that could be expected.

The comparatively low within-test correlations of 0.64 (1969) and 0.67 (1974) between fluency and originality are not dissimilar to the between-test correlations, and this gives some support to maintaining more than just a fluency score. There is a less strong case for keeping the two scores in the Uses and Consequences Tests with within-test correlations in the region of 0.80. Final comments on this question will be better made after the final factor analysis in the last section of this study when divergent tests will be looked at in relation to other variables.

There is some support for the contentions of Yamamoto and Frengel (1966) and Plass *et al* (1974) that the abilities of fluency, flexibility and originality are not clearly related across tests, correlations between fluency scores, for example, being less than correlations between different abilities derived from a single test. To summarize the data and to compare it with other studies total scores (after standardisation) have already been calculated for the individual tests rather than by adding all the fluency or originality scores, and the intercorrelation analysis suggests that this procedure is appropriate. A composite score for the verbal tests is also included where it is thought to be useful, but the correlations between tests are not high enough to suggest that divergent thinking is a unidimensional construct.

For convenience it is often useful to adopt a single score but to do so would be to set up a concept of divergent thinking with a unitary flavour in much the same way as has been the case with intelligence. The separate test scores will therefore be maintained throughout this study though in addition a composite verbal score may be utilized if illuminating. Figural and verbal divergent scores will be kept separate throughout. Although the Circles Test correlates with the verbal tests almost as well as they do between themselves (particularly in the 1969 sample) an earlier factor analysis of the writer (Richards, 1970) showed that the Circles Test has some factorial separation from the verbal tests. As Vernon (1971) observes the Circles Test has often shown different patterns of response from other tests and he would omit it from a divergent test battery. This is slightly reminiscent however of Terman's (1906) deletion of 'imaginative' tests from his I.Q. battery and the very differences shown by the Circles Test makes it valuable in a study of this type. Considerations of its validity and that of the verbal tests, will be dealt with in the next two chapters.

The magnitude of the intercorrelations in the two sets of data provides considerable evidence for convergent validity and is similar to that found in a number of other studies (Hargreaves and Bolton, 1972; Richards, R.L., 1976; Atkinson, 1977). It does not support the findings of some other researchers (Hudson, 1966; Nuttall, 1971) who report low correlations between divergent measures. The overall range of correlations between the three tests, excluding the within-test correlations, is from 0.41 to 0.64 in 1969, and from 0.46 to 0.59 in 1974. In contrast the correlations between the divergent thinking measures and I.Q. are considerably lower, from 0.17 to 0.53 in 1969, and from 0.13 to 0.43 in 1974. In each case the lower correlations occur between the Circles Test and I.Q. The complete intercorrelation table is given in the Appendix.

The intercorrelations between the three divergent tests are, in fact, similar to those reported by Heim (1970) between the verbal/numerical and diagrammatic halves of the AH5 which range from 0.49 to 0.62. Heim's comments that the two parts, though designed to give a total 'operative' I.Q. score, are concerned with somewhat different qualities and should have a significant positive relationship rather than an identical function; could be modified to support the adoption of the present tests into a battery designed to give a single divergent thinking score. Although this is contrary to the writer's position, the three tests used appear to have some potential in this respect. Further evidence will be forthcoming from the investigation of their separate claims to concurrent and construct validity.

9.4 SUMMARY

Three tests of divergent thinking, Circles, Uses, and Consequences were readministered to 139, 15-year-old children, $4\frac{1}{2}$ years after they had originally taken the tests. Coefficients of long-term stability were calculated for the whole sample and for boys and girls separately. These were all significant and were similar in magnitude for both boys and girls.

Stability coefficients for the whole sample ranged from 0.40 to 0.47 for the individual abilities, and from 0.47 to 0.57 for combined scores. Although not exceptionally high, these were considered large enough to indicate that divergent thinking ability shows some definite stability over time. This conclusion was given support by the close agreement shown to exist between the present results and those of other similar studies.

A consideration of levels of performance in 1969 and 1974 showed that divergent thinking ability showed reasonable development over time, and together with the stability in relative ranking, suggested that divergent thinking develops in a similar way to other mental abilities.

A correction for attenuation due to the unreliability of the tests indicated that the results could be potentially even more significant. Assuming, as indicated in the last chapter, that the short-term reliability of the Uses Test is not as great as the other tests, there were some grounds for believing that, over the long term, the divergent thinking ability underlying the Uses Test is more stable than that of the other tests.

Although there was some overlap with general intelligence, the latter was not sufficient to account for the long-term stability of the divergent thinking measures. After the effect of intelligence was removed the stability coefficients were still very substantial, ranging from 0.35 to 0.47 for the individual tests.

Intercorrelations between the divergent thinking tests in 1969 and between the same tests in 1974 showed a very similar pattern, and gave further support to the stability of the underlying construct. On both occasions the level of intercorrelations between the divergent thinking tests was also shown to be greater than between the divergent tests and intelligence.

Overall, the comparison of divergent thinking performance in 1969 and 1974 yields a large amount of evidence in support of a construct of divergent thinking which shows sufficient stability over time to be a significant factor in predicting future behaviour. It also suggests that there are grounds for extrapolating any concurrent validity that may be demonstrated in the next chapter.

CHAPTER TEN

RESULTS OF THE INVESTIGATION OF THE CONCURRENT VALIDITY OF TESTS OF DIVERGENT THINKING

10.1 INTRODUCTION

The basic hypothesis to be investigated in this chapter is that divergent thinking abilities have some predictive validity in relation to a number of criteria of creative behaviour. The criteria to be used have already been described in Chapter 5, and comprise five main measures:

(i) Teacher Rating Scale: 'Indicators of Creative Behaviour'.

(ii) Torrance's Creative Interests Checklist: 'Things You Have Done in your Spare Time'.

(iii) The Golann Creative Motivation Scale: 'Things I Would Like to do Most'.

(iv) 'Interests' questionnaire: including Imaginative, Logical/analytic, and General components.

(v) The Board Game.

These yield seven individual measurements and it is suggested that divergent thinking abilities, if they have relevance to creative activity, should be significantly related to these criteria. This contention put in the form of a null hypothesis for each of the criteria used, will be assessed in the following pages.

Creativity, however, as pointed out in Chapter 2, is not an unidimensional activity but relates to a broad category of intellectual and other variables, and it is possible that some of the divergent abilities will show greater relationships with some criteria than others. The three tests of divergent thinking themselves yield even different ability measures and, without the composite scores, there are consequently 49 correlations between the criteria in the divergent tests. This in itself gives a good chance of finding some significant correlations, so that isolated coefficients of limited size will be treated with some caution.

The validity coefficients are given as basic zero-order coefficients of correlation but an assessment of their practical value is more realistically made if the effects of intelligence are controlled. First-order partial coefficients controlling for intelligence will therefore be presented alongside the basic validity coefficients.

Each of the scales, except for the Board Game, has a good distribution of scores which are consistent with the use of Pearson's product-moment method of correlation. The scale with the smallest range is that of Logical/Analytical Interests, but the scores were well distributed on a nine-point scale from 0 to 8. Guilford and Fruchter (1973) note that there is nothing that demands that Pearson 'r' be computed only with normal distributions, and that even rectangular distributions, with frequencies nearly equal along the range of measurements, would be acceptable. The scores for the Board Game range on a six-point scale from 0 to 5, but the respective frequencies for the whole population, 59, 31, 21, 22, 16 and 12, while being roughly rectangular from o to 5, are heavily weighted in the '0' category, the asymmetry being more marked for girls than for boys. With the resulting lack of symmetry and with so few categories, they are not considered suitable for reliable investigation using Pearson 'r'. The latter may reveal some indications of its possible relationship to the divergent-thinking abilities and the Board Game score will be included in the initial analysis on this basis, but at this stage its statistical analysis for significance will only be made on a tentative basis. It will be investigated in a separate categorical analysis later in this chapter, and this more detailed treatment should also do more justice to its role in the present study.

When considering children's interests and creative behavious it is possible that sex will be a significant variable and the validity coefficients are discussed for each sex separately as well as for the whole population.

10.2 RESULTS AND DISCUSSIONS

10.21 Intercorrelations between the Divergent Tests, and between the Creative Criteria

Intercorrelations, means and standard deviations of all the basic variables, for both sexes separately and for the whole population are given in the Appendix, data relevant to the present discussion being extracted or calculated separately where appropriate.

In discussing the validity coefficients of the different divergent measures with the various criteria, some indication of the relationships within each batch of measures will be illuminating and will affect the way in which their interrelationships are viewed. Before proceeding with this discussion therefore a general picture of the correlations within each battery are given in Tables 23 and 24 respectively.

			1	2	3	4	5	6	7	8	9	10	11	12
Circles	Flu.	1	.	83	64	91	58	47	58	62	42	60	66	29
	Flex.	2		.	63	92	56	37	52	53	41	54	59	23
	Orig.	3			.	85	33	36	38	41	46	50	49	37
Combined Circles		4				.	55	45	55	58	48	61	65	33
Uses	Flu.	5					.	63	91	55	34	52	81	19
	Orig.	6						.	90	50	46	55	82	31
Combined Uses		7							.	59	44	59	90	28
Consequences	Flu.	8								.	49	86	80	28
	Orig.	9									.	86	72	38
Combined Consequences		10										.	88	38
Combined Verbal		11											.	37
I.Q.		12												.

(decimal points are omitted in the table)

Table 23 Correlations between the Divergent Thinking Tests and between the Divergent Tests and I.Q. (whole population n = 161)

(a) Boys n = 77

		1	2	3	4	5	6	7	8
Teacher Ratings	1	.	17	28	09	03	25	43	69
Torrance Checklist	2			18	22	14	15	21	09
Creative Motivation	3				03	07	12	11	17
General Interests	4					81	52	17	05
Imaginative Interests	5						17	03	02
Logical/Analytic Interests	6							21	20
Board Game	7								38
I.Q.	8								

(b) Girls n = 84

		1	2	3	4	5	6	7	8
Teacher Ratings	1		25	35	18	28	14	14	69
Torrance Checklist	2			29	52	51	34	01	18
Creative Motivation	3				35	51	21	24	33
General Interests	4					74	69	05	24
Imaginative Interests	5						26	00	22
Logical/Analytic Interests	6							24	21
Board Game	7								29
I.Q.	8								

(decimal points are omitted in the tables)

Table 24 Correlations between the Creative Criteria and between the Creative Criteria and I.Q.

The general pattern of the correlations in Table 23 is similar to those between the divergent thinking measures in the investigation of long-term stability presented in Chapter 8, and the correlations between the divergent abilities from different tests, ranging from 0.33 to 0.62 are not so high as to suggest that they are measuring identical abilities. The within-test correlations can also be interpreted in this way, with the exception of Circles (fluency) and Circles (flexibility), and are lower than those given in Chapter 8.

In the present sample the correlation between the fluency and flexibility scores for Circles is 0.83, compared with 0.86 and 0.92 for the 11 and 15-year-old populations respectively in the stability investigation, and it consequently appears that the flexibility score is so consistently and highly related to fluency as to be largely superfluous. The correlation of 0.63 between the fluency and originality scores for the Uses Test, however, is lower than those of 0.81 and 0.86 for the populations in Chapter 8; and similarly the intercorrelation of 0.49 for the Consequences Test is considerably lower than the corresponding correlations of 0.78 and 0.78. This gives some support for the modified scoring procedures adopted with the present sample as described in Chapter 6.

With comparatively low intercorrelations even within tests, there is consequently a chance that the different abilities will show different relationships to creative criteria and this is important in this section on concurrent validity.

As noted earlier the creative criteria also cover different types of creative interests and activities and this is borne out by the correlations in Table 24. The only exception is the measure of General Interests which was formed from the same scale as Imaginative and Logical/Analytic Interests and includes both the latter scores. As a result it is highly correlated with both, particularly Imaginative Interests. Its value to the subsequent analysis is therefore in some doubt and its position will be returned to again shortly. There is a low positive relationship between most of the other measures and they are clearly not assessing identical activities.

The chief differences between the correlations for boys and girls centres on the scales for Imaginative and Logical/ Analytic Interests. The former is more clearly related to the other criteria for girls, while the opposite is true for the boys. I.Q. gives an identical correlation with Teacher Ratings for each sex, and otherwise shows slightly stronger relationships for girls. On the Board Game however intelligence is a more significant factor for boys. Apart from Teacher Ratings the creative criteria are not so highly related to intelligence as to be largely a reflection of the latter, and this gives some support to their construct validity.

10.22 Basic Validity Coefficients

The basic validity coefficients indicating the relationship between the predicting variables of divergent thinking and the creative criteria, for the whole population are presented in the following table (Table 25).

		Teacher Ratings	Torrance Checklist	Creative Motivation	General Interests	Imaginative Interests	Logical/Analytic Interests	Board Game
Circles	Flu.	26	34	20	14	14	07	20
	Flex.	14	29	13	10	12	02	19
	Orig.	30	22	17	06	06	19	31
Combined Circles		26	32	18	06	06	10	26
Uses	Flu.	18	29	22	25	19	08	17
	Orig.	31	14	26	13	18	04	16
Combined Uses		27	24	27	21	21	07	18
Consequences	Flu.	26	31	21	23	23	04	-01
	Orig.	36	17	24	15	14	12	13
Combined Consequences		36	28	26	22	22	09	07
Combined Verbal		35	29	29	24	24	09	15
(I.Q.		69	13	26	15	11	18	31)

(decimal points are omitted in the table)

for n = 161 values of r $\geq$ 0.16 are significant $p < 0.05$ } 2-tail test.
values of r $\geq$ 0.21 are significant $p < 0.01$ }

(Correlations significant at 5% are underlined)

Table 25 Validity Coefficients between Divergent Thinking Tests and Criterion Variables

Without including the combined scores, 29 out of the 49 validity coefficients are significant, 18 at 1% and 11 at 5%, adopting a two-tail test. At a less stringent level of 10% a further nine coefficients are significant. Overall this provides a fairly convincing picture of the relevance of divergent thinking abilities to creative activity and provides some real evidence of concurrent validity. The results consequently give weight to studies which have found support for the concurrent validity of divergent thinking tests rather than the negative evidence reported by Kogan and Pankove (1974).

The level of the coefficients is not high however, the significant correlations ranging from 0.16 to 0.36. These compare with the range of significant validity coefficients from 0.32 to 0.48 reported earlier from Torrance (1969) between composite fluency, flexibility, originality and elaboration scores and three criterion of creative achievement obtained from a questionnaire. With only 46 subjects however, the level of significance is similar to that obtained here, though only one out of twelve coefficients did not reach significance in Torrance's study. Combining the divergent data into a single measure Torrance was able to raise the level of validity coefficients, and the resulting values of 0.46, 0.50 and 0.51 compared well with those of 0.42 (girls) and 0.52 (boys), also with combined measures of divergent thinking and creative activities, reported by Cropley (1972).

Vernon (1972) obtained significant coefficients ranging from 0.24 to 0.46 for boys and 0.25 to 0.53 for girls in the study reviewed earlier in Chapter 4. The coefficients were obtained between test scores and a single combined criterion of creativity formed by adding together ten measures including essay ratings, teacher and peer judgements, and leisure-time artistic and scientific activities. The validity of the total divergent thinking (D.T.) battery against the combined creativity criterion was 0.51 for boys and 0.63 for girls respectively. Some of the correlations between the D.T. battery and some of the individual measures however, fell as low as 0.08 for boys and 0.02 for girls. The lowest correlation between a combined divergent test score and the criteria in Table 25 is 0.07 between the Consequences Test and the Board Game. Apart from one slightly negative correlation of -0.01, between Consequences (fluency) and the Board Game, the correlations are all positive and suggest that a combined verbal plus figural divergent thinking score, and a combined creativity criterion would produce results consistent with those of Torrance, Cropley and Vernon.

Dewing (1970) also evaluated the validity of her divergent thinking tests against a combined criterion, and the latter included both Torrance's Checklist and the Golann Creative Motivation Scale. Dewing investigated the validity of the D.T. measures by comparing the performance of a highly divergent group with a 'control group' formed from the remainder of the sample, but she reported a relationship between the groups and the creative criteria significant at the 0.1% level in favour of the divergent group.

In order to compare the present results more directly with these studies using a composite criterion, a 'total' validity measure for figural plus verbal divergent thinking tests, against the total battery of creative activities and interests is computed here using the 'pooling square' method of Thomson (1950) described by Wiseman (1966).

The composite validity coefficients for the figural, verbal and total divergent battery are given in Table 26.

	Boys n = 77	Girls n = 84
Figural D.T.	0.38	0.37
Verbal D.T.	0.42	0.47
Whole Battery D.T.	0.44	0.46

(all correlations significant $p < 0.001$)

Table 26 Validity Coefficients of the Battery of Divergent Thinking Tests with the Battery of Creative Criteria

The validity coefficients of the whole battery are 0.44 for boys and 0.46 for girls. These are very similar in magnitude to those reported by Torrance (1969) and Cropley (1972), and with the greater number of subjects in the present study the coefficients are slightly more significant. Although not as high as those of Vernon (1972), they provide some positive support for his findings.

Like Dewing's (1970) relationship which was evaluated by means of a χ^2 test, the present composite validity coefficients are significant beyond the 0.1% level. Although not directly comparable with Pearson 'r', Dewing also reported a contingency coefficient of 0.39 for 2 d.f. giving an alternative indication of the strength of the relationship.

The derivation of a composite validity coefficient from the results of several D.T. tests and creative activities, while giving a measure of the overall relationship tends to obscure relationships which are present, to a greater or lesser extent, between individual D.T. measures and particular criteria. Although the general level of validity coefficients between individual tests and specific criteria are likely to be lower it is suggested that they will provide a clearer picture of the validity of the tests. Those obtained in the present study will therefore be looked at in more detail shortly.

Putting the overall level of prediction in perspective however, it appears from the present results that, while individual divergent thinking tests appear able to predict a fairly consistent amount of variance in creative criteria the actual proportion of variance is small. The median correlation of the 29 significant coefficients is 0.29, and this for example, predicts only 8.4% of the variation in the criterion, and the highest individual correlation of 0.36 between the originality score for the Consequences Test and Teacher Ratings predicts only 13%. Correcting for attenuation due to unreliability of the predictors and the criteria however would raise this estimate.

Guilford and Fruchter (1973) suggest that many a validity coefficient reported in the literature fails to take into account errors of measurement, and although one may be compelled to deal with fallible tests one should at least allow for unreliability in the criterion itself. The reliability of ratings they point out is characteristically about 0.60, though they also give an instance of teacher ratings of students' creative ability in design as high as 0.82.

Thorndike and Hagan (1969) also warn that the reliability of rating procedures is generally low, though the effectiveness is improved, they suggest, when the rater's response is called for on a structured rating instrument. The latter procedure was adopted in the design of the present Teacher Rating scale and Thorndike and Hagan note that such rating scales, put into the form of a behaviour checklist with a three-point appraisal scale, have been found to yield retest reliabilities between 0.77 and 0.93, and an interscorer reliability of 0.83.

Assuming, to avoid over-correction, that the teacher ratings here have a reliability of 0.85, and adopting the test-retest reliability of 0.60 for Consequences (originality) as given in Table 7 (Chapter 8), the above validity coefficient of 0.36, after correction for attenuation, is raised to 0.51. This increases the amount of predicted variance from 13% to 26%. Applying the same correction to the non-significant correlation of 0.13 between the Consequences (originality) score and the Board Game, assuming that the reliability of the latter is the same as for Teacher Ratings, the validity coefficient is raised to 0.18. Such a correction would bring a number of the lower correlations into the significance category, and bearing in mind the dangers of underestimating the validity of divergent thinking tests because of their unreliability, it can be suggested that the overall pattern of the correlations in Table 25 is indicative of a consistent if low degree of concurrent validity.

The extent to which this degree of validity is likely to be useful depends on whether there are other variables that predict the same variance, and whether any unique variance is of sufficient magnitude to make a significant contribution.

Even after correction for attenuation however, there is certainly a great deal of variance which might be attributable to other variables and an obvious candidate for a source of common variance is I.Q. Looking at the correlations between I.Q. and the criterion variables at the foot of Table 25, four out of the seven correlations are significant, but only that with Teacher Ratings is larger than those achieved by the divergent tests. Apart from the latter, intelligence does not appear likely to swamp the effect of the divergent tests, though whether the correlations imply a common or complementary source of variance awaits to be seen.

Intelligence is however a very substantial predictor of children's creative behaviour as assessed by the Teacher Ratings. The type of behaviour assessed can be seen in the rating scale 'Characteristics of Children's Creative Behaviour' shown in the Appendix, and includes descriptions of the creative child as curious, non-conforming, flexible, willing to try different approaches, unwilling to give up, self-sufficient, original, imaginative and willing to experiment.

The writer has already presented some evidence regarding the use of teacher ratings as criteria of creativity (page and although some doubts were expressed, the writer felt that the alternative views were sufficient to warrant their inclusion. It would consequently be wrong at this stage to attribute their relationship to intelligence as a 'halo effect' due to the more conventional abilities of the children with high I.Q.

The scale might have been better constructed however, if some of the traits have been worded in a positive fashion and others in a negative way so that the distinctness of each trait would be emphasised. Alternately the ratings for each characteristic might have been worded differently, instead of the consistent A to E rating which the teachers were asked to apply to all the traits. Although such modifications would have made the scale less easy to score, they might have encouraged the rater to assess each descriptive statement carefully, and hence have reduced the 'halo' tendency for raters to mark a similar position for all traits according to a general rather than specific assessment of the children's creative behaviour.

The 'halo' effect is therefore likely to be part of the explanation for the high correlation between the Teacher Ratings and Intelligence, but the class teachers, as argued earlier, have an unique opportunity of being closely involved with the children in a range of activities over a considerable period of time, and some considerable weight has to be given to their evidence. The high correlation of 0.69 is consequently viewed as evidence of the concurrent validity of intelligence tests in predicting characteristics of creative behaviour.

It is also noteworthy that all but one of the divergent thinking measures in Table 25 are also significantly related to the Teacher Ratings. Whether this can be explained in terms of a common relationship with intelligence will be investigated shortly.

The correlations between the three scales derived from the Interests questionnaire have already been noted, General Interests being particularly related to Imaginative Interests. A consideration of the validity coefficients with the three scales suggests further that the 'General Interests' measure is superfluous. As noted in Chapter 5, Barker-Lunn (1970) distinguished only two "consistent sub-scales" in the Interests questionnaire, but the writer suggested that seven out of the remaining nineteen items also appeared relevant to creative thinking abilities. These consisted of drawing, collecting stamps, writing a daily diary, doing crossword puzzles, sewing, gardening and dancing, and they were included with the sum of the other two scales into an overall interest measure. Considering the extensive analysis given to the scale by Barker-Lunn, this action of the writer took some liberties with remaining items and it is clear that the result adds very little information to that provided by the two sub-scales. There are more items in the imaginative sub-scale than for logical interests and the overall score is largely a reflection of the score for imaginative interests.

As suggested by Barker-Lunn the imaginative and logical scales show some different patterns of results and this is a further reason for treating an overall score with some suspicion.

The negligible extra value of the overall score is in fact a good indication of the validity of Barker-Lunn's analysis and of the limited value of personal judgements when it comes to the construction of scales of interests and attitudes. There are consequently no grounds for retaining this score as a criterion measure, and although its relationship to the other variables will be looked at again in the factor analysis in the next chapter, it will be omitted as a criterion of creative activity in the following section.

10.23 An Appraisal of the Validity Coefficients for each Divergent Thinking Measure in Conjunction with Sex Differences and the Effect of Intelligence

The basic zero-order validity coefficients for each divergent thinking ability, and the corresponding first-order partial coefficients after controlling for intelligence, are looked at in this section for each criterion measure in turn, beginning with Teacher Ratings in Table 27. Both sets of coefficients are presented for each sex and for the whole population. The combined score for the Circles Test was arrived at by standardising and adding the individual scores for fluency, flexibility and originality, and a similar procedure was used for each verbal test, and to form an overall verbal score.

10.231 Validity Coefficients with Teacher Ratings of Children's Creative Behaviour

As can be seen from Table 27, before controlling for the effects of intelligence, six out of the seven validity coefficients for the individual divergent thinking abilities are significant for the whole population, and so are the coefficients for all four combined scores.

For girls, three out of the seven coefficients were significant, for Circles (fluency), Uses (originality) and Consequences (originality). Another three correlations of 0.21 however were on the borderline of significance at the 5% level, and taken together, provide further evidence of a consistent degree of validity. For boys five coefficients were significant, three of them for the same abilities as for girls, and the other two for Circles (originality) and Consequences (fluency). Both the latter were of borderline

			Teacher Ratings: Boys (n=77)	Girls (n=84)	All (n=161)
Circles Test	Fluency	Basic	$\underline{24}$	$\underline{26}$	$\underline{26}$
		First Order	11	05	09
	Flexibility	Basic	09	17	14
		First Order	04	-09	-02
	Originality	Basic	$\underline{41}$	21	$\underline{30}$
		First Order	19	-05	06
	Combined Score	Basic	$\underline{28}$	$\underline{24}$	$\underline{26}$
		First Order	12	-04	05
Uses Test	Fluency	Basic	11	21	$\underline{18}$
		First Order	01	08	07
	Originality	Basic	$\underline{26}$	$\underline{32}$	$\underline{31}$
		First Order	01	19	13
	Combined Score	Basic	21	$\underline{29}$	$\underline{27}$
		First Order	01	15	11
Consequences Test	Fluency	Basic	$\underline{28}$	21	$\underline{26}$
		First Order	11	01	09
	Originality	Basic	$\underline{38}$	$\underline{34}$	$\underline{36}$
		First Order	17	12	$\underline{16}$
	Combined Score	Basic	$\underline{37}$	$\underline{33}$	$\underline{36}$
		First Order	16	08	15
Combined Verbal Score		Basic	$\underline{35}$	$\underline{34}$	$\underline{35}$
		First Order	10	13	14

(Decimal points are omitted in the table)

Values of r are significant at the 5% and 1% levels as follows:

for n = 161 $r \geq 0.16$ and $r \geq 0.21$ respectively
for n = 84 $r \geq 0.22$ and $r \geq 0.28$ respectively
for n = 77 $r \geq 0.23$ and $r \geq 0.29$ respectively
(2-tail test).

Basic coefficients have n - 2 d.f. and first-order coefficients n - 3 d.f.

Significant coefficients are underlined.

Table 27 Validity Coefficients for the Divergent Thinking Tests with Teacher Ratings, and First-order Partial Coefficients Controlling for Intelligence significance for girls so that overall the pattern with each sex is broadly similar. The originality score for the Circles Test however was a considerably better predictor of creative behaviour for boys rather than girls.

With the common variance due to intelligence removed from the criterion and the divergent thinking measures, the picture changes considerably. Instead of all the validity coefficients for the combined scores for each test being significant, both within each sex and for the whole population, none of the partial validity coefficients for combined scores are significant. For the whole population, only one out of the seven individual divergent thinking measures, Consequences (originality), shows a significant partial correlation with the criterion, and none of the partial validity coefficients are significant for boys and girls separately. It is clear that the evidence for the concurrent validity of the divergent tests shown by the basic validity coefficients is largely due to the effects of intelligence, and that apart from this divergent tests do not predict characteristics of children's creative behaviour as seen by their teachers.

It should be recognised however that, as discussed in Chapter 9, removing the effects of intelligence from both criterion and predictor may remove more common variance than is reasonable, and this is particularly true when, as at present, the I.Q. test is likely to be the most reliable of the tests used. The lower reliabilities of the divergent tests and the criteria, also limit the size of the correlation coefficients which exist between them. A number of the partial validity coefficients, which show some positive residual correlation with the criterion, therefore deserve looking at more closely after correcting for attenuation to allow for the low reliabilities of the tests used. Although such validity coefficients are of little practical predictive value they can contribute evidence of construct validity between the measures under discussion.

With a large number of correlations to choose from, it would not be appropriate simply to correct any of the larger coefficients which appear in the table, but consistent groupings of correlations deserve further attention. The partial validity coefficients for the Circles Test (Boys), the Uses Test (Girls), and the Consequences Test (whole population) show some consistent positive validity and are therefore looked at more closely.

Assuming, as earlier in this chapter, that the reliability of the Teacher Ratings is 0.85, and adopting the test-retest reliabilities of the divergent tests given in Table 7, a correction of the residual validity coefficients for attenuation gives the following indication (Table 28) of the maximum relationship which might exist between the selected divergent abilities and the teachers' assessments of the children's creative behaviour.

Out of the seven measures looked at four remain non-significant, one becomes significant at 1% rather than 5% and two others become significant, one at 5% and the other at 1%. Each of the significant coefficients is for originality and this suggests that efforts to construct more reliable measures of divergent thinking may meet with some moderate success in predicting creative behaviour. The level of the correlations is still low however and effort might better be used in attempting to isolate that element in divergent tests which is responsible for this small relationship.

	Circles Test			Uses Test		Consequences Test	
	Flu.	Flex.	Orig.	Flu.	Orig.	Flu.	Orig.
Partial Validity Coeffs.	0.13	0.05	0.24*	0.12	0.29**	0.11	0.22**
	(0.11)	(0.04)	(0.19)	(0.08)	(0.19)	(0.09)	(0.16*)
	(Boys Only)			(Girls Only)		(Whole Population)	

*Significant at 5%
**Significant at 1%

(Coefficients for the uncorrected data are given in parenthesis)

Table 28 Selected Partial Validity Coefficients Corrected for Attenuation[1]
(Teacher Ratings)

While the Consequences Test had a similar pattern of correlations for boys and girls the originality score for the Circles Test was only a predictor for boys and Uses (originality) only for girls. The difference was most marked in the Circles Test where the partial validity coefficients for girls were negative for flexibility and originality. The function of the Circles Test in this respect will be looked at in relation to the other creative criteria which follow.

---

[1] The correction for attenuation was applied directly to the partial coefficients, and gives only an estimate of the corrected coefficient which would have been obtained by correcting the original correlations before finding the first-order partials. Assuming the I.Q. test to be perfectly reliable this gives an approximation which can be shown by simple algebraic substitution to be always less than or equal to that obtained by correcting the original correlations. An approximation in this direction is convenient for use here. If the partial coefficients had been calculated using the approximate method used by Vernon (1972) the corrections for attenuation would, in fact, have been identical if applied before or after calculating the partials.

10.232 Validity Coefficients with Torrance's Leisure Interests Checklist

The low correlation of 0.13 between I.Q. and Torrance's Checklist of Creative Leisure Interests, taken in conjunction with only moderate correlations between the divergent abilities and I.Q., means that the partial validity coefficients are only slightly less than the basic zero-order coefficients, as can be seen in Table 29. Compared with the I.Q. correlation of 0.13 the overall level of the coefficients indicates that divergent tests have considerably more validity than intelligence tests in predicting children's creative interests.

Six out of the seven basic validity coefficients are significant and only one of these, for Consequences (originality), is not significant after the effect of intelligence is removed. The only non-significant zero-order coefficient is for Uses (originality) but this is significant for girls alone. Consequences (originality) is also significant for girls and not for boys, but the reverse is true for Circles (originality). Once again the figural test appears to be more strongly associated with boys and the verbal tests with girls, though Circles (fluency) is an exception. All of the basic validity coefficients for combined scores, except for the Uses Test for boys, are significant, and they all remain significant after the effects of intelligence are removed.

In contrast to the very low partial validity coefficients reported in relation to Teacher Ratings, the present partial coefficients remain quite substantial and give some definite evidence for concurrent validity. Unlike the small coefficients which did exist for Teacher Ratings, the present coefficients are highest for fluency rather than originality scores.

Test	Score	Coefficient	Torrance Checklist Boys (n=77)	Girls (n=84)	All (n=161)
Circles Test	Fluency	Basic	31	41	34
		First Order	30	38	31
	Flexibility	Basic	30	33	29
		First Order	30	29	27
	Originality	Basic	24	19	22
		First Order	23	13	19
	Combined Score	Basic	31	34	32
		First Order	30	30	29
Uses Test	Fluency	Basic	29	37	29
		First Order	28	34	28
	Originality	Basic	05	27	14
		First Order	02	23	11
	Combined Score	Basic	19	35	24
		First Order	18	32	22
Consequences Test	Fluency	Basic	32	41	31
		First Order	31	38	29
	Originality	Basic	12	24	17
		First Order	10	19	13
	Combined Score	Basic	25	39	28
		First Order	23	35	25
Combined Verbal Score		Basic	26	40	29
		First Order	25	36	26

(Decimal points are omitted in the table)

Values of r are significant at the 5% and 1% levels as follows:

for n = 161, $r \geq 0.16$ and $r \geq 0.21$ respectively
for n = 84, $r \geq 0.22$ and $r \geq 0.28$ respectively
for n = 77, $r \geq 0.23$ and $r \geq 0.29$ respectively
(2-tail test).

Basic coefficients have n - 2 d.f. and first-order coefficients n - 3 d.f.

Significant coefficients are underlined.

Table 29 Validity Coefficients for the Divergent Thinking Tests with Creative Leisure Interests, and First-order Partial Coefficients Controlling for Intelligence

This points to an association between the number of interests and activities children pursue in their leisure time and the fluency of their ideas on divergent tests. In asking children to indicate interests and activities from a prepared list there is a danger however that some children will tick the items with less discrimination and evaluation than others. As noted when describing the test earlier, there is no means of assessing the level at which an interest may function, though the writer discussed their completed checklist with some of the children.

Utilizing some of the teachers' judgements about their pupils' general abilities it seemed likely that some of the children had indicated interests which they were not likely to pursue in any great depth. Although this was evidently the case with some pupils who had ticked a large number of activities, they showed an interest and enthusiasm for the activities which the writer considered generally justified their inclusion.

Different children interpreted the activities differently and this is part of what appears to be assessed by the test. Instead of being an objective list of activities, they become relevant to the subjects who take the test and it is their interpretation which is recorded. Some children indicated that they had 'made up or organised a play or sketch' (No.12) on the basis of sketches at Boy Scout meetings or parties, while others who had also taken part in such activities, were less ready to claim that they had made them up or organised them. 'Kept a science scrapbook' (No.47) was interpreted quite reasonably by one less able girl in terms of her scrapbook of "space-travel and things like that" but 'science' was more narrowly interpreted by others. One boy noted that he had 'Organised or helped to organise a club' (No.55) because he belonged to a 'gang' and helped to organise it, and another claimed to have 'Made an electric motor' (No.29) because he had taken one out of an old toy car and built it into a boat.

A large number of low scorers could also have been made to realise that they had also done something to justify ticking certain items, but their reluctance to do so is obviously part of what the test is measuring. Total scores on the test ranged from 12 to 82 out of 95 listed activities.

Although some researchers (e.g. Eysenck, 1967; Ausubel, 1968; Shouksmith, 1970) might view uninhibited responses to the interests questionnaire as evidence of the same illconsidered dilletantism of which they have criticised responses to divergent tests, the writer is inclined to view the list of interests in a more positive light.

High scorers appear to be those who are most able to recognise something that they had done as fitting into the description of the activity, while low scorers tend to interpret the activities more narrowly and literally. The checklist may consequently be functioning more as a test of 'creative imagination' than as an inventory of creative activities, but it is suggested that in either case it has some validity as a criterion instrument. Its significant correlations with the divergent tests then give, in turn, some concurrent validity to the latter.

With basic validity coefficients in excess of 0.40 and partial validity coefficients as high as 0.38 there is a good deal of evidence for concurrent validity. A consideration of construct validity by correcting for attenuation in order to explore the possibility of a significant underlying relationship is therefore not necessary, but there is obviously some scope for increasing the present validity coefficients by making the tests and the criterion more reliable.

10.233 Validity Coefficients with Golann's Creative Motivation Scale

The validity coefficients with creative motivation are given in Table 30. All the basic zero-order validity coefficients, except for Circles (flexibility), are significant for the population as a whole, four of them at a 1% level, though the overall level of the coefficients is low. Within the sexes the coefficients are quite substantially positive, but are significant only for girls in the two verbal tests.

After controlling for intelligence the criterion still has significant residual correlations with the verbal tests for the population as a whole, with three out of four of the individual abilities and with each of the combined measures. None of the partial validity coefficients for the Circles Test are significant however, neither within sexes nor for the whole population. The reduction is due chiefly to the effects of intelligence being more marked for girls than for boys and this trend also existed for Teacher Ratings and Leisure Interests. Once again the verbal tests also tend to be better predictors of the creative criterion for girls rather than boys.

Only the combined verbal score is significant for boys alone, though there is a consistent pattern of relationships with the other abilities bordering on the 10% level of significance. With consistent correlations of this type summed

			Creative Motivation		
			Boys (n=77)	Girls (n=84)	All (n=161)
Circles Test	Fluency	Basic	19	21	20
		First Order	15	11	13
	Flexibility	Basic	11	14	13
		First Order	10	03	07
	Originality	Basic	20	15	17
		First Order	14	04	08
	Combined Score	Basic	18	18	18
		First Order	14	07	10
Uses Test	Fluency	Basic	18	25	22
		First Order	16	19	18
	Originality	Basic	18	32	26
		First Order	13	25	20
	Combined Score	Basic	20	32	27
		First Order	16	25	21
Consequences Test	Fluency	Basic	14	27	21
		First Order	10	19	14
	Originality	Basic	22	26	24
		First Order	17	16	16
	Combined Score	Basic	20	31	26
		First Order	15	21	18
Combined Verbal Score		Basic	24	34	29
		First Order	19	25	22

(Decimal points are omitted in the table)
Values of r are significant at the 5% and 1% levels as follows:
for n = 161, $r \geq 0.16$ and $r \geq 0.21$ respectively
for n = 84, $r \geq 0.22$ and $r \geq 0.28$ respectively
for n = 77, $r \geq 0.23$ and $r \geq 0.29$ respectively
(2-tail test).
Basic coefficients have n - 2 d.f. and first-order coefficients n - 3 d.f.

Significant coefficients are underlined

Table 30 Validity Coefficients for the Divergent Tests with Golann's Creative Motivation Scale, and First-order Partial Coefficients Controlling for Intelligence scores usually give a better indication of any underlying relationship, and the differences between the sexes are also small enough to rely on coefficients from the whole population.

There are no marked differences between the correlations obtained with individual abilities within tests, except for the low correlations with Circles (flexibility). Overall the whole pattern is one of a positive but rather weak relationship which gives some slight evidence of concurrent validity.

The results are not unlike those obtained by Golann (1962) in validating the scale against a revision of the Barron-Welsh Art Scale. With eight groups of children of 11 and 13 years of age, the size of the group varying between 15 and 55, Golann found seven positive correlations ranging from 0.16 to 0.41, four of which were significant at 5% or better. He concluded that while the magnitude of the relationship was not large it showed a stable positive relationship between the B-W scale and the creative motivation scores.

Golann did not correct his correlations for attenuation, though he quoted a test-retest reliability coefficient of 0.66 for the 11-year-old groups over a period of three weeks. Applying a correction for attenuation on the basis of this reliability coefficient and those given in Table 7 for the divergent tests the validity coefficients obtained for the whole population are increased as shown in Table 31.

(n=161)	Circles Test			Uses Test		Consequences Test	
	Flu.	Flex.	Orig.	Flu.	Orig.	Flu.	Orig.
Zero-order Coeffs.	0.27** (0.20*)	0.19* (0.13)	0.24** (0.17*)	0.37** (0.22**)	0.45** (0.26**)	0.30** (0.21**)	0.38** (0.24**)
Partial Coeffs.	0.18* (0.13)	0.10 (0.07)	0.11 (0.08)	0.30** (0.18*)	0.35** (0.20*)	0.20* (0.14)	0.25** (0.16*)

*Significant at 5%
**Significant at 1%

(Validity coefficients for the uncorrected data are in parenthesis)

<u>Table 31 Validity Coefficients with the Creative Motivation Scale (corrected for attenuation)</u>

With this correction, the one zero-order validity coefficient not initially significant becomes significant at 5%, and the coefficients rise as high as 0.45 for Uses (originality). The level of the correlations now indicates quite a high potential level of association between creative motivation and divergent tests. Only two of the partial validity coefficients remain non-significant, for the Circles Test (flexibility and originality). As a result of the lower reliability of the Uses Test the partial validity coefficient for originality rises from 0.20 to 0.35.

Golann (1962) reports a higher test-retest reliability for his scale with older children (r = 0.86 for 13-year-olds), and the potential coefficients indicated by the above correction for attenuation is therefore not entirely academic.

Although the present coefficients are not as high as those between the B-W Art Scale and divergent abilities (0.44 to 0.49) given by Lang and Ryba (1976), the latter experiment was conducted with an able population of university students mainly studying courses in creative arts. Considering the age and range of ability of the children in the present study and the possible over-compensation by removing the effects of intelligence, the small but consistent relationship between Golann's scale and the verbal divergent thinking abilities in particular, provides some positive evidence for concurrent validity even if the level of the relationship is small.

10.234 Validity Coefficients with Imaginative Interests

The basic zero-order validity coefficients and the first-order partial coefficients controlling for intelligence are given in Table 32. There are clear differences in this set of correlations between the verbal and figural tests and between sexes.

None of the validity coefficients for the Circles Test are significant though there is a low positive overall relationship. Both verbal tests however are significantly correlated with the imaginative interests criterion, for the population as a whole, the combined scores being significant at 1% before controlling for intelligence and at 5% afterwards. Only Consequences (originality) is not significant for the individual verbal scores. The overall effect however is due very largely to the relationship for girls alone. In fact of all the individual coefficients for boys, only that for Circles (originality) is in any way appreciable, and considering its isolated nature it could have been a chance association.

By contrast all the zero-order coefficients except Consequences (originality) are significant for girls though none of them reach the 1% level. After controlling for intelligence they become non-significant at 5% though five out of

			Imaginative Interests		
			Boys (n=77)	Girls (n=84)	All (n=161)
Circles Test	Fluency	Basic	09	10	14
		First Order	09	03	12
	Flexibility	Basic	00	14	12
		First Order	-01	07	09
	Originality	Basic	15	08	06
		First Order	16	00	02
	Combined Score	Basic	09	12	12
		First Order	09	04	09
Uses Test	Fluency	Basic	-02	$\underline{25}$	$\underline{19}$
		First Order	-02	21	$\underline{18}$
	Originality	Basic	00	$\underline{22}$	$\underline{18}$
		First Order	-01	17	$\underline{16}$
	Combined Score	Basic	-01	$\underline{26}$	$\underline{21}$
		First Order	-02	21	$\underline{19}$
Consequences Test	Fluency	Basic	05	$\underline{23}$	$\underline{23}$
		First Order	04	18	$\underline{21}$
	Originality	Basic	09	14	14
		First Order	09	06	11
	Combined Score	Basic	08	$\underline{22}$	$\underline{22}$
		First Order	07	15	$\underline{19}$
Combined Verbal Score		Basic	04	$\underline{26}$	$\underline{24}$
		First Order	03	20	$\underline{21}$

(Decimal points are omitted in the table)
Values of r are significant at the 5% and 1% levels as follows:
for n = 161, $r \geq 0.16$ and $r \geq 0.21$ respectively
for n = 84, $r \geq 0.22$ and $r \geq 0.28$ respectively (2-tail test).
for n = 77, $r \geq 0.23$ and $r \geq 0.29$ respectively
Basic coefficients have n - 2 d.f. and first-order coefficients n - 3 d.f.
Significant coefficients are underlined.

Table 32 Validity Coefficients for the Divergent Tests with Imaginative Interests, and First-order Partial Coefficients Controlling for Intelligence the seven coefficients remain significant at 10%. Correcting for attenuation would in turn increase the significance of the latter and the overall level of the coefficients, though the coefficients for boys would remain non-significant for all the tests. None of the partial coefficients for the Circles Test reach significance after correction though the corrected basic validity coefficient for fluency becomes significant at 5%.

Barker Lunn (1970) pointed out that the Interests Scale showed differences between the scores of boys and girls, imaginative interests appealing more to girls than boys. A similar picture emerges here with girls having a mean score of 8.30 (s.d. 2.35) and boys 6.00 (s.d. 2.62). The within-sex analysis however allows possible relationships with divergent abilities to emerge separately, and it is clear that divergent thinking abilities have some validity in predicting imaginative interests for girls, and that this relationship is not dependent on intelligence to more than a very small extent.

Barker Lunn suggested however that children of above average intelligence seemed to choose imaginative activities of a more demanding nature, though girls of less than average ability tended to have a higher overall score. As discussed in relation to Torrance's checklist, the relationship between divergent thinking abilities and children's self-report questionnaires of creative interests, could be regarded as a common facility for making indiscriminate associations or as a deficiency of self-criticism and evaluation.

On the other hand children of high intelligence do not have a monopoly of interest in creative activities, and children of primary school age are generally willing to express their enthusiasm for an activity even if they are not able to fully assess their level of accomplishment. The scale asked for the degree of enjoyment of various activities rather than any specific level of performance and it seems that children, irrespective of their level of intelligence, show different capacities to involve themselves creatively in a variety of activities and that this is reflected to a small but significant extent by tests of divergent thinking. Gardner (1950) found in a study of the interests of 10-year-old children, that those from classrooms in which the curriculum was 'interest orientated' showed a higher level of activity in pursuing their interests than children from more 'traditional' classrooms; and it is consequently likely that the degree of children's enthusiasm for creative activities can also be developed with appropriate teaching.

As in Wallach and Wing's (1969) study of talented college students, the present results indicate slightly more association between the fluency scores and creative activities than for originality scores. This is true for both Torrance's checklist and the Imaginative interests, though there are also some significant validity coefficients for the originality scores. The latter are less open to objection on the grounds of common irrelevant fluency, though in spite of the objections the writer would also claim some considerable validity from the fluency coefficients.

10.235 Validity Coefficients with Logical/Analytic Interests

Both zero-order coefficients and first-order partial

Test	Measure	Coefficient	Logical/Analytic Interests: Boys (n=77)	Girls (n=84)	All (n=161)
Circles Test	Fluency	Basic	10	18	07
		First Order	05	12	02
	Flexibility	Basic	-03	16	02
		First Order	-05	10	-03
	Originality	Basic	14	18	<u>19</u>
		First Order	07	12	13
	Combined Score	Basic	08	19	10
		First Order	02	13	04
Uses Test	Fluency	Basic	21	16	08
		First Order	18	12	05
	Originality	Basic	05	20	04
		First Order	-03	15	-02
	Combined Score	Basic	15	20	07
		First Order	09	15	02
Consequences Test	Fluency	Basic	12	20	04
		First Order	06	15	-02
	Originality	Basic	09	<u>24</u>	12
		First Order	02	18	05
	Combined Score	Basic	12	<u>26</u>	09
		First Order	05	20	02
Combined Verbal Score		Basic	16	<u>25</u>	09
		First Order	09	19	02

(Decimal points are omitted in the table)
Values of r are significant at the 5% and 1% levels as follows:
for n = 161, $r \geq 0.16$ and $r \geq 0.21$ respectively
for n = 84, $r \geq 0.22$ and $r \geq 0.28$ respectively (2-tail test).
for n = 77, $r \geq 0.23$ and $r \geq 0.29$ respectively
Basic coefficients have n - 2 d.f. and first-order coefficients n - 3 d.f.
Significant coefficients are underlined.

Table 33 Validity Coefficients for the Divergent Tests with Logical/Analytic Interests, and First-order Partial Coefficients Controlling for Intelligence coefficients are presented in Table 33, though only one of the coefficients, the basic coefficient for Circles (originality), is significant ($p < 0.05$) for the whole population. Unlike most of the coefficients with other criteria however, most of the coefficients with logical/analytic interests are less significant for the whole population than for boys and girls separately. This reflects their different levels of preference for these interests, the mean score for boys being 5.60 (s.d. 1.70), and for girls 4.02 (s.d. 1.73). Within sexes there are some significant relationships for girls but not for boys. The trend of coefficients for boys is generally positive, with that for the fluency score for Uses being significant at 10% for the zero-order coefficient, but like the coefficient between Circles (originality) and imaginative interests in the last section this is an isolated coefficient which might have occurred by chance.

On the other hand it appears that girls who have logical and analytic interests do tend to have higher divergent thinking scores. All the basic coefficients are consistently positive, those for Consequences (originality), Consequences (combined score) and Verbal Tests (combined score) being significant at 5%. The combined scores for Circles and Uses are both significant at 10½, and with substantial, though non-significant, coefficients remaining for all the scores after the effects of intelligence are removed, the evidence for the girls provides some concurrent validity.

Even for girls however, the level of the coefficients is not high, though a correction for attenuation would increase their magnitude to a potentially more useful level. Assuming a reliability of 0.85 for the interests scale, the validity coefficient for Consequences (originality), for example, would increase from 0.24 to 0.34.

As well as considerations based on the significance and magnitude of the validity coefficients the weight that one gives to the results depends to a large extent on the value that one can place on the validity of the criterion itself. Like its imaginative counterpart on the Barker Lunn scale it is the children's own expression of interest in the activities, rather than any assessment of the level of their performance, which is measured by the list of logical/analytic interests, and the views expressed in the last session regarding the imaginative scale apply similarly to this criterion.

Although the writer has suggested that considerable weight can be given to the evidence of concurrent validity already considered, the objections that have been acknowledged regarding criteria which depend on children's responses to self-report inventories are not applicable to the next criterion, and it is suggested that the validity coefficients derived in relation to the following practical criterion deserve special recognition.

10.236 Concurrent Validity of Divergent Thinking Tests in relation to the Board Game

The nature of the data derived for the Board Game has already been outlined in Chapter 5, and the distribution of the results described earlier in this chapter (page 320). In view of the asymmetry of the distribution validity coefficients derived using Pearson 'r' are not likely to be very accurate though it is not possible to estimate exactly what error is involved in such correlations. Coarse grouping tends to result in correlations which are underestimates of a true underlying relationship, though the effects of skew distributions are not so consistently predicted. Wylie (1976) has investigated the effects of both on the Pearson product moment coefficient however and produced some relevant findings.

Taking a normal bivariate distribution with known correlation, he grouped the data to form 24 categories in each variable, then gradually reduced the number of categories, incorporating different degrees of asymmetry, and computing a Pearson 'r' for each new set of data. As an example, he found that with one variable reduced to only two categories and with extreme skew in opposite directions, the theoretical correlation of 0.40 was nearly halved to 0.21. With the same variable reduced to six categories however (as in the present Board Game experiment) the Pearson 'r' was a fairly accurate 0.393.

For all the varying degrees of skewness and numbers of categories the value of Pearson 'r' ranged from 0.37 to 0.41 provided one variable had at least twelve categories and the other at least six. Corrections for coarse grouping narrowed this range to 0.39 to 0.41, but a larger number of correlations were then slightly overestimated. To reduce the possibility of the latter no correction for grouping is made to the data in the present study. The data supplied by Wylie however gives some idea of the accuracy of the Pearson coefficients for the Board Game, the error very likely being considerably less than 10%. If precise significance levels were required this degree of error would be unacceptable, but with only one coarsely grouped variable and with some caution in interpreting the general levels of significance, correlations between the divergent thinking measures and the Board Game should give some useful information about their relationship. This will be followed by a categorical analysis in the next section. The 'tentative' validity coefficients of the divergent tests with the Board Game are therefore presented in Table 34, with coefficients of possible significance underlined. As in the case of logical/analytic interests boys score more highly than girls on this variable, the mean score for boys being 2.22 (s.d. 1.77) and for girls 1.11 (s.d. 1.43). Some variation in performance between the sexes was expected, but the extent of this difference was not revealed by the pilot investigations with a comparatively small number of children. The within-sex correlations may therefore be more revealing, and the subsequent analysis will pay particular attention to the validity of the Board Game for each sex separately.

With cautious reference to the significance levels of 5% and 1% as indications of the significance of the validity coefficients in question, there is considerable evidence of some concurrent validity particularly for boys. Considering the more even distribution of performance by boys their validity coefficients are likely to be more accurate than those for girls, and the values of 0.32, 0.38, 0.23 and 0.36 for the combined scores for Circles, Uses, Consequences and Combined Verbal, respectively, are all significant, the three larger coefficients at well beyond 1%. Apart from the Consequences score these remain significant after the removal of the effects of intelligence, and suggest that divergent tests

			Board Game		
			Boys (n=77)	Girls (n=84)	All (n=161)
Circles Test	Fluency	Basic	28	25	20
		First Order	21	17	12
	Flexibility	Basic	22	29	19
		First Order	20	21	13
	Originality	Basic	35	23	31
		First Order	24	15	22
	Combined Score	Basic	32	29	26
		First Order	24	20	18
Uses Test	Fluency	Basic	39	13	17
		First Order	37	07	12
	Originality	Basic	27	22	16
		First Order	15	16	07
	Combined Score	Basic	38	20	18
		First Order	30	13	11
Consequences Test	Fluency	Basic	22	-05	-01
		First Order	13	-15	-10
	Originality	Basic	18	17	13
		First Order	04	07	02
	Combined Score	Basic	23	07	07
		First Order	10	-05	-05
Combined Verbal Score		Basic	36	15	15
		First Order	24	05	04

(Decimal points are omitted in the table)

Tentatively, values of r are significant at the 5% and 1% levels as follows:

for n = 161, $r \geq 0.16$ and $r \geq 0.21$ respectively
for n = 84, rr ≥ 0.22 and $r \geq 0.28$ respectively
for n = 77, $r \geq 0.23$ and $r \geq 0.29$ respectively
(2-tail test).

Basic coefficients have n - 2 d.f. and first-order coefficients n - 3 d.f.

Coefficients of likely significance are underlined.

Table 34 'Tentative' Validity Coefficients for the Divergent Tests with the Board Game and First-order Partial Coefficients Controlling for Intelligence can play a significant role in evaluating the type of ability needed for success on experimental tasks such as the Board Game.

There is no clear pattern to suggest that one of the attributes of fluency, flexibility or originality is particularly important, all three figuring prominently in one test or another. For boys the validity coefficient for originality is greatest in the Circles Test, that for fluency in the Uses Test and the combined score in the Consequences Test. For girls the validity of the originality measures is more marked, and it is the only ability significant in two tests, namely the Circles Test and the Uses Test.

The flexibility measure which was included in the Circles Test with the Board Game in mind, is significally related to performance on the latter for girls and for the population as a whole, but is only of borderline significance for boys. For girls it is in fact the highest individual validity coefficient. However, as discussed in connection with Guilford's Model of the Intellect in Chapter 3 and in the scoring procedures in Chapter 6, credit for originality also demands a considerable degree of flexible thinking, with the added demand for quality of response.

The amount of weight to give to the tentative correlation coefficients in Table 34 is not yet clear however and an alternative method of analysis is needed to throw further light on this question.

10.24 Categorical Analysis of the Validity of the Divergent Tests in Relation to the Board Game

This analysis will be carried out for a selection of the divergent measures and the results generalised to cover the remainder of the validity coefficients in Table 34 not directly checked by categorical analysis. As boys and girls show differing patterns of results they will be treated separately in this section.

Two measures are chosen for investigation within each sex, one figural and one verbal. There is no special significance about the choice, but only measures whose coefficients appear significant in Table 34 are selected. For the verbal tests these are those with the highest tentative validity coefficients, namely Uses (fluency) for boys and Uses (originality) for girls. To represent the figural measures the combined score for Circles is used for both boys and girls.

The children were ranked in each sex separately according to performance on each of these divergent thinking measures and three groups formed for each measure. These were comprised of boys or girls in the top 30% (High group), middle 40% (Average group) and bottom 30% (Low group) on each of the respective distributions. The performance of the children in each of these groups on the Board Game is illustrated in Tables 35, 36, 37 and 38.

When tied ranks occurred at boundary points the individuals concerned were allocated to whichever group needed the greater proportion of the tied group.

The data in each table is analysed using the χ^2 test. The significance of χ^2 depends only upon the degrees of freedom in the table, no assumptions being made as to the form of the distribution of the variables classified (Garrett, 1966; Guilford and Fruchter, 1973). The null hypothesis for each table is that there is no relationship between divergent thinking ability and performance on the Board Game.

	Board Game Responses						Totals
	α	β_1	β_2	β/γ	γ_1	γ_2	
High D.T.	6	3	8	3	1	2	23
Average D.T.	3	3	6	4	8	7	31
Low D.T.	3	2	3	3	2	10	23
Totals	12	8	17	10	11	19	77

Table 35 Children's Performance on the Board Game at Different Levels of Divergent Ability

(Circles Test (Combined Score). Boys, n = 77)

χ^2 Analysis (Table 35)

The data is reduced to a 3 x 3 table for the calculation of χ^2 in order to make the expected frequency in each cell exceed 5 (Nesbitt, 1966).

Performance on the Board Game was therefore reduced to three classifications, High, Medium and Low, by adding the α and β_1 categories, the β_2 and β/γ categories and the γ_1 and γ_2 categories respectively.

As a result $\chi^2 = 9.588$ for 4 d.f.

This is significant ($0.01 < p < 0.05$) and the null hypothesis of no association between performance on the Board Game and Divergent Thinking is rejected.

As an indication of the degree of the relationship, the coefficient of contingency, C, is 0.33.

	Board Game Responses						Totals
	α	β_1	β_2	β/γ	γ_1	γ_2	
High D.T.	1	2	3	6	6	7	25
Average D.T.	2	2	2	4	8	16	34
Low D.T.	1	0	0	1	6	17	25
Totals	4	4	5	11	20	40	84

Table 36 Children's Performance on the Board Game at Different Levels of Divergent Ability

(Circles Test (Combined Score). Girls, n = 84)

χ^2 Analysis (Table 36)

For the calculation of χ^2 the data is reduced to a 3 x 2 table with two classifications for the Board Game, 'success' (α, β_1, β_2, and β/γ) and 'failure' (γ_1 and γ_2).

$\chi^2 = 9.82$ for 2 d.f. ($0.005 < p < 0.01$).

The null hypothesis of no association is rejected.

As an indication of the degree of the relationship $C = 0.32$.

	Board Game Responses						
	α	β_1	β_2	β/γ	γ_1	γ_2	Totals
High D.T.	7	2	5	5	1	3	23
Average D.T.	3	5	10	4	5	5	32
Low D.T.	2	1	2	1	5	11	22
Totals	12	8	17	10	11	19	77

Table 37 Children's Performance on the Board Game at Different Levels of Divergent Ability

(Uses Test (Fluency). Boys, n = 77)

χ^2 Analysis (Table 37)

The data was reduced to a 3 x 3 table as for Table 35.

$\chi^2 = 16.55$ for 4 d.f. $(0.001 < p < 0.005)$

The null hypothesis of no association is rejected.

As an indication of the degree of the relationship $C = 0.42$.

	Board Game Responses						Totals
	α	β_1	β_2	β/γ	γ_1	γ_2	
High D.T.	1	2	1	5	6	9	24
Average D.T.	2	2	4	4	8	14	34
Low D.T.	1	0	0	2	6	17	26
Totals	4	4	5	11	20	40	84

Table 38 Children's Performance on the Board Game at Different Levels of Divergent Ability

(Uses Test (Originality). Girls, n = 84)

χ^2 Analysis (Table 38)

The data was reduced to a 3 x 2 table as for Table 36.

$\chi^2 = 5.39$ for 2 d.f. $(0.05 < p < 0.1)$

The null hypothesis of no association is in doubt.

As an indication of some relationship between the variables C = 0.25.

The results of the χ^2 analyses shown in Tables 35 to 38 lead to the same conclusions as the Pearson 'r' validity coefficients in Table 34. The validity coefficients for Uses (fluency) and Circles (combined score) both for boys, which are significant at 1% in Table 34 are of a similar, but slightly lower level of significance by χ^2 analysis at 1% and 5% respectively. The relationship for Circles (combined score) for girls is slightly more significant by χ^2 analysis than for the correlation coefficient ($p < 0.01$), and Uses (originality) for girls is rather less significant, $p < 0.07$, compared to $p = 0.05$ exactly for the correlation coefficient.

Overall the levels of significance are slightly reduced but considering that non-parametric tests are less 'powerful' than parametric tests with the same data the categorical analysis provides convincing support for the general level of validity coefficients in Table 34, in spite of the skewness and coarse grouping in the Board Game distribution.

As an illustration of the lack of power of the non-parametric tests it is interesting to note that in the case of Uses (originality) for girls, three girls with tied divergent thinking scores appeared on the borderline between the top and middle groups and were all successful at the Board Game with scores of 5, 3 and 1 respectively. In choosing the top 30%, one of these children should have been placed in the top group, but with tied scores all three were placed in the middle category as originally decided. Had they been placed in the top category the value of χ^2 would have been increased markedly, and the value actually obtained, significant at approximately 7% is therefore regarded as of borderline significance comparable to that obtained for 'r'.

Parametric tests use more of the detail from the data than is used by non-parametric tests and the fact that this group of children had comparatively high divergent scores would have been utilized in the former method of analysis.

Coefficients of contingency, C, are also given as an alternative expression of the value of χ^2 and give some indication of the magnitude of the relationship. Under certain conditions, of fine grouping, underlying normality in the categorised variables and large samples, C is a good estimate of 'r', but in the present case these conditions are not fully realised. C is therefore included only as a general indication of the likely underlying value of 'r'. There is a general level of agreement between C and the values of 'r' in Table 34, C values of 0.33, 0.32, 0.42, and 0.25 corresponding to values of 'r' of 0.32, 0.29, 0.39 and 0.22 respectively.

Reduction of the divergent thinking scores into three categories loses a considerable amount of information, but in addition to making the data convenient for χ^2 analysis it has the advantage of being able to illustrate the nature of individuals' performances on the Board Game. Looking at Table 36 for example it illustrates that although the high divergent thinking girls, as assessed by the Circles Test, are not guaranteed successful performance on the Board Game, the higher their performance the greater their likelihood of success. In the high D.T. category 12 children were successful and 13 unsuccessful while in the middle category these figures were 10 and 24 respectively, and in the low category only two were successful out of 25. While the latter category was also below average in I.Q., five of the girls had I.Q.'s greater than 100, and the girl with an 'α' performance in the

Board Game had an I.Q. of only 88.

Although divergent ability is not a criterion for success on the Board Game for individuals such as the latter, further study of individuals in depth might reveal other characteristics that might enable them to gain success, and the Board Game could provide a useful technique to help in such an investigation. Evidence regarding other abilities that might be relevant to success on the Board Game will be looked at in the next chapter when the general relationships between the variables are examined.

In particular the characteristics of the above girl suggest that her extraverted nature might be one factor contributing to her success in the Board Game, though it did not enhance her performances on the divergent tests which were consistently low. She was above average in extraversion and rated as very lively and seldom shy of her teacher. She was also far less neurotic than the average girl and had slightly above average self-image regarding school work, and self-concept in general. She had relatively few interests however, as assessed by all the interest scales and she was rated very low by her teacher for characteristics of creative behaviour.

She appears more of a doer than a thinker and a 'scanner' rather than a 'focusser' (Bruner *et al*, 1956). Her spontaneous actions resulted in success at the Board Game though her I.Q. and divergent performance and the 'average' ratings for school work suggest that her powers of concentration might be limited. The danger with scanners is that they rely a great deal on trial and error action and find sequential work difficult. As a result they tend to be relegated in the more conventional levels of school work. Further research utilizing the Board Game might help reveal some of the characteristics of scanners and focussers and provide further information regarding the general trend established above relating divergent thinking ability to successful performance.

With reference to the classification of boys' performance in Table 37 it can be seen that 19 out of the 23 high divergent thinkers have some success on the Board Game as opposed to only 6 out of 23 in the low group. Of the 12 α-solutions by boys, 7 were obtained by the high group.

As can be seen from the correlations in Table 25, and by the significant partial validity coefficients in Tables 26 to 34, the divergent thinking measures are not highly correlated with I.Q., and this indicates that the individuals in a high divergent group are not those who could have otherwise been selected, on the basis of intelligence. The fact that children in the high divergent groups on various divergent thinking measures tend to perform significantly better on the Board Game than those in lower groups gives considerable weight to the concurrent validity of the tests. This claim of divergent thinking tests to possess incremental validity over and above that of I.Q. is looked at further in the next session.

10.25 Relative Performance of Convergers and Divergers on the Board Game

A closer look at the children who appear in the top 30% on the intelligence test or divergent thinking, or both, can provide a clearer indication of the positive value of divergent thinking ability in comparison with I.Q. For the purpose of this illustration the Board Game will be used as the criterion measure and the Uses (fluency) score for boys alone will be used to select the divergent thinking group. The Uses (fluency) score is the best predictor of success at the Board Game for boys, and its correlation with I.Q. is only moderate ($r = 0.39$), so that it should be possible to demonstrate its power of prediction over and above that achieved by the I.Q. test. The same general principle holds for each of the other divergent measures and criteria and this point will be returned to shortly.

The performance of the two groups at the Board Game is shown in Table 39.

	Board Game Responses						
	α	β_1	β_2	β/γ	γ_1	γ_2	Total
High I.Q.	5	5	7	2	3	2	24
High D.T. (Uses, fluency)	7	2	5	5	1	3	23

Table 39 Performance at the Board Game of Boys in the Top 30% on I.Q. or Divergent Thinking

That both I.Q. and divergent measures are clearly relevant to success at the Board Game is shown by the table, and more generally, by their correlations of 0.38 and 0.39 respectively with Board Game performance. The extent to which the same children are registering success in both groups however is not clear from Table 39 and a consideration of the following Genn diagram (Figure 8) is more illuminating. It illustrates the Board Game performance of the two groups of boys selected by I.Q. and D.T. (divergent thinking) respectively.

The two loops represent the top 30% of boys on each measure, and the numbers inside the loops give the Board Game score of the children within each group. The overlapping area represents those children who appear in both groups - the all-rounders, and the remaining areas are those of the convergers (those in the top 30% on I.Q. but not in the top 30% on D.T.) and divergers (defined conversely).

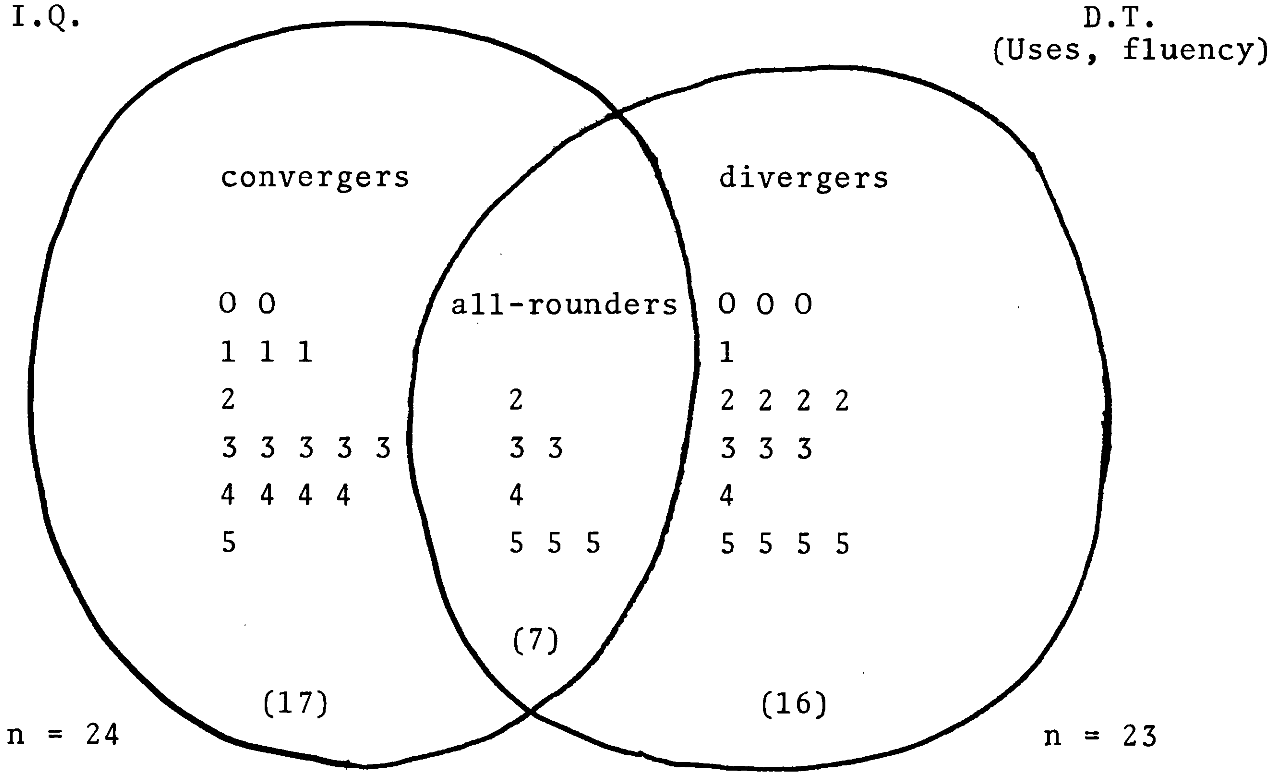


Fig.8. Performance on the Board Game of Boys in the Top 30% on I.Q. or Divergent Thinking (or both)

It is evident from Figure 8 that the validity of divergent thinking in predicting boys performance on the Board Game is not simply a duplication of that of intelligence. In contrast to Hasan and Butcher (1966) who found it difficult to establish groups of convergers and divergers due to the substantial overlap between the measures of I.Q. and D.T., two such groups are readily distinguishable in the above distribution, though the present cut-off points are set at 30% rather than 20%. Overall the Board Game performance of the convergers and divergers is very similar, though the divergers included more of the very best solvers. Both groups included a number of boys who failed to achieve a successful outcome but this was a smaller proportion than in the sample as a whole. In the latter 30 out of 77 boys did not gain a successful solution, being in categories γ_1 or γ_2, though only 4 out of 23 divergers and only 5 out of 24 convergers were in these categories. In the group of all-rounders however none of the boys failed to achieve some success, 3 out of the 7 achieving α-solutions.

In general terms the average performance of the all-rounders (mean 3.86, s.d. 113) is considerably better than that of both the convergers (mean 2.71, s.d. 1.56) and the divergers (mean 2.63 s.d. 1.76). The differences are significant at just over and just under the 5% level ($t = 2.054$ d.f. 22, and $t = 2.085$, d.f. 21) respectively using the t-test for small samples.

If the 23 boys in the top 30% on the Circles Test (combined score), whose performance on the Board Game is shown in Table 35, were added into Figure 8, 5 new children would be added, 11 would appear in common with those already chosen by I.Q., and 11 in common with D.T. (Uses, fluency). Four boys would appear common to all three groups, being in the top 30% on all three measures. Each of these four obtain successful solutions, three achieving α-solutions and one a β_2-solution.

Divergent thinking tests clearly have considerable validity in predicting success at the Board Game, and function, to a substantial degree, independently of I.Q. in choosing potential solvers. As a result the combination of I.Q. and divergent thinking tests provided an even more marked effect in predicting the Board Game performance and this suggests that combined measures could provide considerable information about children's abilities. Although the Venn diagram analysis has been conducted for boys only, comparable effects would occur whenever there are significant validity coefficients in Tables 35-38.

10.26 An Illustration of the Potential of Divergent Thinking Tests and I.Q. in Forming a Multiple Validity Coefficient

In statistical terms the maximum combined effects of more than one measure in predicting another can be presented in terms of their multiple correlation. As a more general expression of their combined effect than was illustrated by the Venn d agram, the effect of both intelligence and divergent thinking scores in predicting boys performance on the Board Game can therefore be indicated by their multiple correlation, R, with the Board Game. For practical use the scores on both tests would be combined according to a multiple regression equation, which weights the scores so as to achieve maximum prediction. Provided that the two independent variables do not correlate highly there can be a considerable gain in validity by combining them in this way.

For example the multiple validity coefficient for a combined predictor of I.Q. and Uses (fluency) in relation to performance on the Board Game for boys is 0.51. This compares with individual correlations with the Board Game of 0.38 and 0.39 respectively, and a partial correlation of 0.37 between

Uses (fluency) and the Board Game with the effect of the first variable (intelligence) removed.

For girls the correlation between I.Q. and the Board Game is 0.29, and it is also 0.29 between the Board Game and divergent thinking (assessed by Circles (combined score) as in the χ^2 analysis). The latter validity coefficient is reduced to 0.20 after the effect of intelligence is removed but the subsequent multiple R, combining I.Q. and D.T. is raised to 0.35. The borderline validity coefficient of 0.22 for Uses (originality) for girls is reduced to 0.16 after controlling for intelligence but even this small residual effect increases the validity of I.Q. alone (0.29) to give a multiple R of 0.33.

In the first of these examples there is quite a marked increase in predicted variance of 13%, but in the other two cases the increase is only 4% and 2½% respectively. Although considerable evidence has therefore been presented in this chapter to show that divergent thinking tests possess statistically significant amounts of concurrent validity the actual magnitude of the predictions has not been high.

Some further prediction might be gained however by adding other tests to the calculation of multiple correlation. Guilford and Fruchter (1973) point out that the addition of tests to a predictive battery which have intercorrelations of about 0.60 or more adds little to the overall predictive effect. For example, assuming validity coefficients of 0.30 for each test in the battery and intercorrelations ot 0.60, the multiple R for two tests is 0.34 but for 20 similar tests is only 0.38. Although some of the correlations between the individual divergent thinking abilities in the present battery are as high as 0.62, most of them are in the region of 0.50 as can be seen in Table 23. Some, in fact, particularly between figural originality and the verbal measures, are as low as 0.33 and, provided they have significant validity coefficients, these could make valuable contributions to extending the multiple correlation of the battery.

The additional effect of considering Circles (combined score) in the prediction of boys performance on the Board Game has already been mentioned in the Venn diagram analysis, and it is a suitable measure to add to the multiple correlation approach to predictive validity. It has correlations of only 0.28 and 0.46 with I.Q. and Uses (fluency) and a validity coefficient of 0.32. Calculating the correlations r_{12} between the Board Game (B.G.) and I.Q., $r_{13.2}$ between B.G. and Uses with I.Q. removed, and $r_{14.23}$ between B.G. and Circles with both I.Q. and Uses removed, the multiple correlation $R_{1(234)}$ between the Board Game and a combination of the other three variables can be determined. This gives a value of $R = 0.57$, and shows a substantial increase on that of 0.51 for I.Q. and a verbal D.T. measure alone.

The addition of a second divergent test contrasting with the first therefore accounts for an additional 6½% of the variance in the criterion, and the multiple R predicts approximately 32½% of the criterion variance. This is a substantial improvement on the prediction of 14½% by I.Q. alone. Not all of the divergent tests are as effective predictors as those chosen for this example, but it points to the considerable potential of divergent thinking tests in this respect.

10.3 SUMMARY AND CONCLUSIONS

10.31 General Evidence for Concurrent Validity

The main conclusion from the evidence presented is that divergent thinking abilities have some positive validity in relation to a number of criteria of creative behaviour.

Considering the six creative criteria finally selected, and the seven divergent thinking abilities, three from the Circles Test, two from the Uses Test and two from the Consequences Test, a substantial proportion of the validity coefficients are significant. For the population as a whole 27 out of the 42 coefficients are significant, 16 at 1% level and 11 at 5%. A further 6 would be significant at 10%, or if a 1-tail test had been adopted.

Creative activity in one sphere does not necessarily imply creativity in another and it was not intended to establish an unidimensional battery of tests of creative behaviour. Apart from one coefficient however, there was a low positive relationship between the measures for both boys and girls (Table 24) and a combined criterion score allows an overall validity index to be calculated. For a whole battery of divergent thinking tests the resulting validity coefficients are 0.44 for boys and 0.46 for girls, both highly significant ($p < 0.001$). Separate validity coefficients for the figural and verbal measures of divergent thinking are also significant at this level, the relationship being stronger for verbal tests, particularly for girls.

The results consequently give support to studies which have found positive evidence for the concurrent validity of divergent thinking tests, and the general level of the relationship is also similar to a number of other studies.

As found by Vernon (1972) the overall divergent battery appears to be a better predictor for girls than for boys, though in this study this is due chiefly to the better validity of the verbal divergent thinking measures for girls.

10.32 The Effect of Intelligence on the Validity Coefficients of the Divergent Tests

First-order partial validity coefficients controlling for the effect of intelligence demonstrated that part of the validity of the divergent tests was equally well accounted for by intelligence. Out of the 27 validity coefficients which were significant for the whole population however, 13 remained significant at the 5% level or greater, and substantial positive correlations remained in many of the others. Considering the possible over-compensation by removing I.Q. effects, it was suggested that the distinct divergent thinking effect was still of importance, and correction for attenuation for the greater unreliability in the divergent tests than in I.Q. (see Tables 28 and 31), reinstated a number of the coefficients in the significant category.

These results compare well with those of Milgram and Milgram (1976) who in the study referred to in Chapter 4, looked at the concurrent validity of a summed divergent thinking score based on fluency measures alone, in relation to nine areas of creative activity as assessed from a self-report inventory given to 145 high-school students. Partialling out intelligence they found significant validity coefficients in two out of the nine areas for boys and in two areas for girls. Two further coefficients were significant at 10%, and an overall correlation with a combined score for creative performance was significant at 1% for both boys and girls separately. Using the combined figural and combined verbal scores from the present study, four out of the twelve partial validity coefficients were significant for boys, and three for girls. A further three coefficients were on the border-line significance at 5%.

Milgram and Milgram concluded that their results showed that students reporting high creative performance were characterised by a higher level of creative thinking - though they noted that the relationship was not high. A similar conclusion, on a slightly firmer basis, can be made for the results in this study.

The effect of intelligence was considerably more marked in some criteria than others and a clearer picture of its effect is gained by considering the individual criteria.

With the exception of Teacher Ratings I.Q. was no more highly related to the creative criteria than were the divergent variables, as can be seen from Table 25. As a result removal of I.Q. had a greater effect on validity coefficients with Teacher Ratings than on the other coefficients. Only one of the six significant validity coefficients with Teacher Ratings remained significant after I.Q. was removed, though in contrast, of the six significant validity coefficients with Torrance's Checklist five partial coefficients remained significant, 4 at the 1% level. A number of features of these criteria were discussed to explain these results, and it is important to accept that I.Q. and divergent thinking can both be related to various criteria, in different ways and to different degrees.

This is shown clearly in the analysis of the result of the Board Game in which I.Q. and divergent thinking abilities showed similar levels of correlation but by no means identified the same individuals. It is suggested that I.Q. and divergent thinking abilities can therefore play complementary roles in identifying creative talent. As an illustration the multiple correlation between a combination of I.Q. and divergent thinking (Uses, fluency), and the Board Game for boys was 0.51, as opposed to individual validities of 0.38 and 0.39 for I.Q. and D.T. respectively.

Correlations between the different divergent thinking measures (Table 23) show that divergent abilities, even for the same tests, though quite highly correlated, are by no means identical, and further prediction can be gained for a battery of tests by considering their optimal combination rather than by simply finding a total divergent thinking score. As an example, when the combined score for Circles was added to the above predictor of success on the Board Game the multiple correlation was raised to 0.57, the combination of divergent abilities contributing an extra 18% of the variance, compared to 14½% by I.Q. alone. A good deal of the evidence for concurrent validity therefore remains, over and above the effect of intelligence.

10.33 Sex Differences and Validity Coefficients with Individual Criteria

Some sex differences in creative activities and interests remarked on by a number of writers (e.g. Barker Lunn, 1970; Vernon, 1972) were confirmed in the present analysis, and these existed not only in terms of some differences in level of performance, but in the pattern of validity correlations with the divergent tests. Although the validity coefficients were generally positive for both sexes the magnitude of some of the correlations varied considerably and the effect of intelligence was usually more marked for girls than for boys.

As shown in Table 26 the overall validity coefficients were slightly higher for girls than for boys, but this is not true of the figural score which was a slightly more valid predictor for boys. The verbal measures on the other hand, possibly reflecting the more consistent effect of intelligence for girls, was considerably higher for girls.

In comparison with the 27 coefficients out of 42 significant to the population as a whole, 14 are significant to boys alone (8 at 1%) and 21 for girls (7 at 1%).

The differences were most marked on the Imaginative Interests scale, none of the divergent tests having more than a nominal degree of validity for boys, but both verbal tests having significant coefficients for girls. This was also true to a lesser extent on the Golann scale, all the verbal scores showing significant validity coefficients for girls but not for boys. The reverse was true for the Board Game, both combined verbal scores being significant predictors for boys but not for girls. Validity coefficients with Teacher Ratings and Torrance's Checklist, and to a lesser extent Logical/Analytic Interests showed less sex differences than the other criteria.

It appears, as might have been expected, that divergent tests function best as predictors of creative activities that are most relevant to the experiences of the subjects. If self-report inventories are considered by the subject to be lacking in relevance to his own experience or situation then his replies are not likely to do justice to his real creative interests. External ratings, provided the rater has sufficient experience of the children and the rating scale is sufficiently wide, seem able to overcome this difficulty, but the Board Game, which was designed as the major criterion, suffers slightly in this respect.

Although some researches (e.g. Sharples, 1969) have suggested that "general assumptions about girls' subjects and boys' activities" are not as well founded as one might think, the result of the present study suggest that even in the present era of unisex, batteries of creative activities and interests might serve as better criteria if they were designed more specifically with boys and girls interests in mind. As with the problems of reliability discussed in Chapter 8, motivation has to be sustained over both the initial test and the criterion test (or retest) in order to gain the maximum predictive validity, and this might have to take boys' and girls' varying preferences into account.

This aspect should not be overstated however as a considerable degree of concurrent validity extends across both criteria and sex differences. Validity coefficients are strongest when the subjects bring the same abilities to bear on both criteria and predictors and motivation is just one of the common variables. The present results provide evidence of a good deal of association between divergent abilities and the creative criteria, even when the association is stronger in one sex rather than another, and there are a number of validity coefficients which assume greater significance in the whole population than with either sex alone.

As discussed earlier it should also be noted that apart from Teacher Ratings and the Board Game the criteria of creative activities carried no evaluation for quality of response. Milgram and Milgram (1976) give some support to the use of self-report inventories scored according to quantity of creative interests, finding a high correlation between their quantity index and a quality measure based on details supplied with the initial information. Their finding however is likely to be more true of their able high school pupils than of the 11-year-olds in this study.

10.34 <u>The Board Game</u>

The Board Game experiment was given particular attention, partly because of the nature of the data, and partly because of the greater weight that the writer suggests should be given to this criterion.

As pointed out when the individual validity coefficients were considered only the Teacher Ratings and the Board Game carry credit for quality of response. Teacher Ratings do so implicitly and subjectively and, as shown, rely to a very large extent on abilities covered by the conventional intelligence test. The Board Game, on the other hand, was designed to involve thinking abilities that called for some hard analytical thinking in "trying to penetrate, to realise and to trace out the main relations between form and task", Wertheimer (1949, 1961), fluency and originality in thinking of unusual ideas and uses for objects, and the flexibility to try them out in different ways.

The writer has already discussed the theoretical background and the practical application of the task in Chapters

3 and 5 respectively, and on this basis the results, particularly for boys are claimed as important evidence of concurrent validity for divergent thinking tests.

This is especially true of the Circles and Uses tests which both gained substantially significant validity coefficients for both boys and girls, as high as 0.39, Uses (fluency), for boys and 0.29, Circles (flexibility and combined score), for girls. These results were confirmed and clarified by a χ^2 analysis.

The Consequences Test was a significant predictor for boys having substantial validity coefficients for the individual abilities and a significant combined coefficient. For girls however Consequences was much less successful, providing a positive correlation for originality but one of the few negative coefficients for fluency. This suggests that adding scores from several different tests is not always appropriate and a far better predictive effect is gained by considering the multiple correlation of the individual variables with a criterion. Combining scores for I.Q., Uses (fluency) and Circles (combined score) in this way achieves a multiple correlation of 0.57 between these three measures and the Board Game. This provides a very substantial multiple validity coefficient and illustrates the potential of divergent thinking tests in this respect.

10.35 Individual Variations in the Validity of the Divergent Tests

Apart from the function of the flexibility score, no clear pattern emerges regarding the differing validities of the separate divergent abilities within tests. Validity coefficients for flexibility however were closely related to fluency coefficients, on all tests and for both sexes, with the fluency score generally being the better predictor of the two. There is consequently little evidence to support the use of separate flexibility scores, and taken together with the correlation of 0.83 with fluency reported in Table 23, the present findings support the views of Wallach (1970) and Hargreaves and Bolton (1972) that the flexibility score is not likely to supply any additional information to that given by the fluency score.

The relationship between fluency and originality is more equivocal. Considering the 18 instances in which they appeared for each sex (3 tests and 6 criteria), 36 pairs of validity coefficients were reported altogether. On 11 occasions the coefficients were significant for both, on 5 occasions they were significant for fluency only and on 5 occasions for originality and not fluency. In the 10 cases showing differences, most of these were slight, the non-significant correlations being substantially positive. In 4 cases however the differences were more marked. On Torrance's Checklist the verbal fluency scores for boys were considerably better predictors than the corresponding originality scores, 0.29 compared to .05 for Uses, and 0.32 compared to 0.12 for Consequences, while on Teacher Ratings boys originality scores were better predictors on Circles and Uses than the corresponding fluency score. For the Board Game the originality score on the Uses Test was a considerably better predictor than the fluency score for the girls, and though neither coefficient was significant a similar pattern was present on the Consequences Test.

There is a suggestion that fluency scores are more related to self-report inventories, and that originality scores are a better predictor when quality of creative performance is assessed. For this reason the writer would support the continued use of originality scores for assessing divergent abilities. Fluency scores have a comparable level of validity however, and for some exercises the time saved by omitting the originality score may be considered worthwhile.

Although there is some slight evidence of these abilities being related across tests, the stronger relationship appears to be with the tests themselves, as suggested by Wallach (1970) and Plass *et al* (1974). This question will be looked at again in the factor analysis in the next chapter.

A more positive conclusion can be made in respect of the value of the Circles Test in assessing divergent thinking abilities. Previous opinions regarding its inclusion in a divergent battery have been varied, and it has showed some factorial separation from other divergent tests in a previous study of the writer (Richards, 1970). The latter was taken as a positive indication of its role in a battery of divergent tests, though Vernon (1971) suggests that its discrepancies with the results of other divergent tests makes it a poor member of a divergent battery. While Vernon therefore suggests that it should be omitted from a divergent thinking battery, Hargreaves and Bolton (1972), as a result of research into selecting 'creativity' tests for use in research, recommend its inclusion.

The results of the present study support the latter view, particularly if verbal and figural tests are combined as multiple predictors rather than by simply adding their scores.

Validity coefficients for the Circles Test were more evenly distributed between the sexes, apart from Imaginative Interests it had one or the other of its scores significantly related to each criterion. It was the only test significant for both boys and girls in predicting success in the Board Game. It was less successful than the verbal tests however in predicting Imaginative Interests and the creative activities on the Golann Scale, both which had a considerable emphasis on creative writing activities.

None of the tests were more than moderately related to intelligence, the highest correlation with I.Q. for the population as a whole being that of 0.38 for Consequences (originality). Considering its lower reliability reported in Chapter 8, the validity coefficients for the Uses Test assume greater significance than those for the other tests if corrected for attenuation, and give some positive evidence for its position, as described by Hudson (1966) as almost the conventional divergent thinking test.

Without the flexibility score for Circles, each test has scores for two abilities which, with six criteria, gave rise to 36 validity coefficients in all. Of the 25 which were significant for the whole population (see Table 25), 9 were for the Circles Test, 9 for Uses and 7 for Consequences. Within the separate sexes, 6 of the validity coefficients for Circles Test scores, excluding flexibility, were significant for boys and 5 for girls. For the Uses Test 4 were significant for boys and 8 for girls, while for the Consequences Test the figures were 3 and 7 respectively. Vernon (1972) reported higher validity coefficients for his battery for girls and Cropley (1972) for boys. Overall the present battery gives a similar pattern to that of Vernon, though the figural measures from the Circles Test are slightly better predictors for boys. As suggested in Section 10.33 it is not always possible to provide a criterion which is free of bias towards one sex or the other, and it is suggested that the possible predominance of either girls or boys will therefore depend on the constitution of both the testing battery and the criterion measures.

With some care in regard to the motivational effects which, as suggested in Chapter 8, might have reduced the reliability of the Uses Test, the writer would recommend the Circles and Uses Tests for future research, as an effective complementary pair of tests, one figural and one verbal.

10.36 General Appraisal of the Evidence Presented

Wallach and Kogan (1966) examining their evidence for effects of divergent thinking ability (creativity) and intelligence on other variables of achievement and personality tested approximately 200 separate hypotheses. They found 42 relationships which they regarded as significant, accepting a significant level of 10% and sometimes even beyond this. With as many as 20 (10%) likely to be 'significant' by chance alone they nevertheless interpreted the evidence as being sufficient to have important implications for education. Some reviewers (e.g. Shulman, 1966) commended their interpretations, others (e.g. Cronbach, 1968) regarded their attempt to draw out implications and applications as premature. This variation in views illustrates the difficulty of assessing the weight that should be given to research results. While some reviewers might place more store by the experimental design and the power of the statistical analysis, others might give more weight to a sound theoretical basis and tentative generalisations.

In establishing the present evidence for validity the writer has acknowledged that there is likely to be some variation in the amount of validity one can claim for the different criteria. While some misgivings might be felt about the value of self-report inventories, the writer has argued that, even though they do not assess the quality of children's responses they still provide evidence of creative activity. Special emphasis however is placed on the Board Game, and with its theoretical background and practical nature, it adds a more realistic dimension to the type of data arrived at from the results of self-report inventories and questionnaires. As noted earlier the writer feels that the significance of the divergent tests in predicting the children's level of performance on this criterion is particularly important evidence for their concurrent validity. The other criteria add up to a more conventional measure and the overall validity coefficients are of a similar magnitude to those found by other researchers.

Evaluating the evidence statistically, the overall validity coefficients are significant beyond the 0.1% level. Excluding Circles (flexibility), 25 out of the 36 separate coefficients for the individual divergent thinking abilities are also significant ($p < 0.05$). The level of significance chosen, a two-tail test at 5%, is not lenient, and considering the unreliabilities of both tests and criteria, the relationships are sufficient to reject the null hypothesis of no association between tests and criteria with considerable confidence.

While statistically significant however the magnitude of the relationships is not high enough to suggest that divergent thinking abilities play more than a small part in predicting creative activity. The median value of the 25 significant validity coefficients, 0.29, predicts only 8.4% of the variance in a particular criterion, though the composite divergent score predicts 19.4% of the total creative criterion score for boys, and 21% for girls.

Both Anastasi (1968) and Cronbach (1970) point out that even tests with validity coefficients as low as 0.20 can make a valuable contribution to a battery of tests, especially if they are not highly correlated with other tests in the battery, and it is in this light that the level of concurrent validity established in the present study can best be interpreted. There has been a consistent residual effect after controlling for intelligence, and this suggests that divergent tests can provide an 'incremental' factor (Sechrest, 1963; Lewis, 1974) in recognising creative potential.

Had convergent abilities, as assessed by I.Q., showed a relationship with the creative criteria to a far greater extent than that of the divergent abilities, the proportion of variance contributed by the latter might have been comparatively unimportant. As it was however, the relationship of I.Q. to the criteria, apart from Teacher Rating, was of a similar magnitude to those achieved by the divergent thinking measures. This might have been expected in the self-report inventories and questionnaires where quality of performance was not assessed, but it is particularly significant in the case of the Board Game, with its demands on children's constructive, analytic and imaginative thinking ability.

In the latter a multiple prediction from I.Q. and two divergent thinking measures, Uses (fluency) and Circles (combined score) accounted for 32½% of the variance in boys performance on the Board Game, and a very appreciable proportion, 18%, was contributed by the D.T. measures.

There are, nevertheless, no grounds for equating high divergent thinking abilities with high performance on any of the creative criteria, including the Board Game, and such abilities are able to predict only a limited amount of the variance in children's creative activities. This proportion however is significant and relatively consistent, and it is suggested that the results give considerable weight to establishing that divergent thinking abilities are worth fostering and developing in relation to creative activity. It may be that alternative and easier ways will be found of assessing this divergent aspect of children's thinking ability, but in the meantime the evidence suggests that divergent thinking tests can play a part in recognising such ability.

As Anastasi (1968) points out, any concurrent validity which is established also provides additional evidence of construct validity for the underlying psychological trait, and although the actual magnitude of the present concurrent validities may, in general, be limited to only a moderate incremental effect in practical applications, they also provide support for a model of the intellect such as that of Guilford (1956, 1967) which includes a category of divergent thinking abilities.

Further information on the construct validity of the divergent tests will be provided in the next chapter, in a factorial investigation of their relationships with a range of other variables.

CHAPTER ELEVEN

RESULTS OF THE FACTORIAL INVESTIGATION OF CONSTRUCT VALIDITY

11.1 INTRODUCTION

This investigation is designed as an exploration of the relationships between divergent thinking tests and a range of other variables. In all scores on 31 variables were derived from the experimental procedures as described in Chapter 5. These consist of the seven separate divergent thinking scores, three for Circles, two for Uses and two for Consequences; four measures of convergent thinking, namely two I.Q. tests and end-of-year assessments by the teacher for English and Mathematics; seven attitude scores covering attitude to school and school work, academic self-image and self-concept; the seven creative criteria used in the investigation of concurrent validity, five measures of personality comprising the Extraversion, Neuroticism and Lie scales of the Junior Eysenck Personality Inventory and 'shy' and 'lively' ratings from the children's class teacher; and a general assessment of the children's pleasurability in class, again given by the respective class teacher.

Placing the latter variable in a 'personality' category for the present, there are consequently five general categories, divergent thinking, convergent thinking, attitudes, personality and creative criteria. Having several tests purporting to represent common abilities make the data amenable to factor analysis, in which tests having common variance help to establish recognisable factors. With this in mind the second I.Q. test, which was not used in the previous sections, and the General Interests score from the Barker-Lunn Interests questionnaire have both been retained in the present analysis.

As a pre-requisite for factor analysis the data has to fulfil the normal conditions for the calculation of Pearson product-moment coefficients, and with the exception of the Board Game these are broadly satisfied by all the other variables. The teacher's assessments of end-of-year performance in English and Mathematics, and the ratings for shyness, liveliness and pleasurability however, were on only a five point scale, but each of the distributions was approximately normal and the resulting coefficients of correlation with other variables should not be gross underestimates of their true correlations. Considering the ratification of the Pearson 'r' validity coefficients for the Board Game in the last chapter, it too is deemed suitable for inclusion in the present analysis, the factor structure being unlikely to suffer from such small variations in isolated variables.

To study the intercorrelations without the aid of factor analysis would mean that over 900 separate correlation coefficients would have to be considered, or rather, double this number, as each sex is to be looked at separately. Factor analysis however enables a 'parsimonious reinterpretation' of the data to be made, a comparatively small number of factors being calculated to account for common relationships in the original set of variables. There is no single 'best' system of factor analysis however and different pictures of the same data are produced by different techniques. The use of a particular system depends, as Fruchter (1954) points out, on the purposes and theoretical approach of the individual factor analysis, and this is amply illustrated as noted in Chapter 3 by different approaches to the study of intelligence (Vernon, 1961; Guilford, 1967; Cattell, 1971).

As both Vernon and Cattell point out however methods of multiple factor analysis have developed a great deal since Spearman's early investigations, and with a common second order factor underlying the primaries extracted using oblique axes, some reconciliation has been possible between the opposing theories of either one general ability or distinct primaries. Cronbach (1970) suggests that the important thing is that different methods of looking at certain psychological traits, like different methods of mapping Greenland, can be constructive and illuminative, without denying the underlying existence of the phenomenon in question. Cattell (1971) nevertheless maintains that the most superior method from the psychologist's point of view is that of rotation of oblique axes to Thurstone's simple structure. This has the advantage of being uniquely determinable and hence not subject to bias on the part of the experimenter, but it results in factors which are themselves intercorrelated.

By contrast an objective mathematical solution can also be achieved by extracting successive linear combinations of the variables, the first combination having the maximum amount of variance, the second as great a proportion of the residual variance as possible, and so on. This direct 'principal axes' method of factor analysis fits in well with a 'general ability' theory of intelligence, as the first factor claims the major share of common variance. It has been widely used to develop hierarchical models of the intellect (Vernon, 1961), but Guilford (1967) suggests that it is more of a statistical artefact than an illustration of the natural order of things.

As something of a compromise the use of a varimax rotation (Kaiser, 1958) of principal axes "is a precisely defined method which indeed approximates to orthogonal simple structure" (Harman, 1967). Thurstone's recommendations for simple structure, summarised by Harman, include having as many zeros in each column of the factor matrix as there are common factors, several variables whose entries vanish in one column but not in the other for every pair of columns in the factor matrix, and at least one zero in each row. While, as Harman notes, these have compelling intuitive value, they lack the precision necessary for mathematical computation. The varimax method however, with an emphasis on maximising and minimising the entries in each column, provides one method of approximating to simple structure, and has become the procedure most often used today (Cronbach, 1970).

The method is obtainable in factor analysis packages at most computing centres, and was employed by the writer via the Computing Unit at the University of Bath using the SPSS programs (Nie *et al*, 1975).

The varimax rotation is applied to an initial principal axes solution and there are two options open for the latter. Following Harman's (1967) nomenclature the initial factor matrix can be either a principal-component solution if unities are placed in the diagonals of the correlation matrix or a principal-factor solution if estimates of communalities are used instead.

As noted in Chapter 1 (1.34) total test variance (σ_T^2) can be divided into two components, true variance (σ_t^2) and error variance (σ_e^2), so that $\sigma_T^2 = \sigma_t^2 + \sigma_e^2$. Some of the true variance however is likely to be specific to the test itself and a more general expression for the purposes of factor analysis is that $\sigma_T^2 = \sigma_c^2 + \sigma_s^2 + \sigma_e^2$ where σ_c^2 is common variance shared with other tests and σ_s^2 is variance specific to the test itself.

Placing unities in the diagonal of the correlation matrix in effect assumes that all the test variance in common variance and this total variance is distributed amongst the factors extracted. The factors therefore include both common variance and some unknown proportion of unique variance. The latter becomes greater as later factors are extracted, and Cattell (1952) suggests that the proportion becomes so great as to swamp the common variance. An estimate of the common variance or communalities of each test, however, if placed instead of unities in the diagonal, automatically builds into the factor model an allowance for unique variance. This is likely to give a truer picture of the underlying relationships between the tests and the resulting factors are known as 'inferred' rather than 'direct' factors.

There is no agreed upon method of calculating communalities though they cannot be greater than one or less than the squared multiple correlation between a variable and all others in the set. Employing the latter as a first estimate the PA2 program for a principal-factor solution uses an iterative process which continues until the best estimates of communality have been found. This is the program adopted here, the final varimax solution being a rotation of the initial principal-factor analysis.

In spite of the sophistication of the mathematical techniques, common factors, as Burt (1940) emphasises, are only statistical abstractions, not concrete entities, and factor analysis is more suitable for creating hypothesis than testing them. Eysenck (1953) however argues that when simple structure principles are applied it also has a role to play in investigations of a hypothetico-deductive nature. Commenting on orthogonal methods of solution in particular, Cattell (1971) nevertheless emphasises that the initial choice of tests can be made to influence the resulting factor matrix in predetermined ways, and he notes that Guilford "makes the initial choice of tests fit a subjective academic framework rather than a naturalistic sampling of existing behaviour".

Factors should not therefore be reified as eternal verities (Fruchter, 1954), but they serve to represent the underlying sources of variation common to a given set of scores in a way which makes it easier to make psychological interpretations about the relationships between the variables. It is important to realise that conclusions are therefore reached via interpretations of the data, and as Guilford (1950) points out, the degree of confidence that can be given to them depends upon the compellingness of the factor structure and the repeated verification of a result, rather than on any statistical test of significance.

The absence of the latter is one of the reasons for McNemar's somewhat exaggerated attacks on factor analysis as a tool of 'Psychological Regress' (McNemar, 1951, 1964). Warnings about the fallibility of the technique however, are salutary when there is a danger that with the widespread provision of computer facilities it could become too 'convenient' a technique with, as McNemar (1964) points out, bigger but not necessarily better factor-analytical studies.

Factor analysis is normally applied in the hope that the first few factors will account for most of the variance so that the remaining factors become of little importance. When all the variance is to be distributed in a principal-components solution there are as many factors as there are variables, unless any variable is an exact linear combination of some of the others. Only a relatively small number however are usually worth investigating, though the decision regarding the number to extract is a matter for the investigator. In this study Kaiser's (1958) criterion is used, only factors with eigenvalues greater than 1.0 being retained. This ensures that only factors accounting for at least a proportion of the total variance equal to that of a single variable, i.e. 3.2%, will be extracted. This condition is applied before the removal of the unique variance, and with communalities in the diagonal the actual proportion of variance allocated can be less than 3.2%.

There are a number of criteria for estimating the significance of factor loadings (Harman, 1967; Burt, 1952; Rippe, 1953) but they are not directly applicable to rotated factor solutions, particularly when only non-unique variance is distributed. Though he compiles a table of standard errors of factor loadings, Harman acknowledges that they are not entirely reliable. Butcher (1969) observes that "No very satisfactory answer appears to have been found to the problem of determining the statistical significance of a rotated factor loading". He adopts an arbitrary figure of 0.35 to distinguish high loadings which, with 70 variables and a population of 1,000 he considers likely to be a conservative estimate of significance.

In view of the uncertainty surrounding the assessment of significance in factorial work Child (1970) suggests that it would be safer to adopt a stringent rather than lenient level, and suggests that provided the sample is greater than 50, loadings above 0.30, accounting for 9% of the variance in that variable, are likely to be significant at quite a rigorous level. Alternatively, he points out that one can interpret the significance of loadings in the same way as Pearson product-moment coefficients, though in this case he suggests one should adopt the 1% level rather than 5%. A single level of significance based only on the sample size, however, makes no allowance for the number of variables or the factor under consideration. Using the Burt-Banks formula (Burt 1952), Child compiles significance levels of factor loadings for different numbers of variables and factors, demonstrating that the acceptable value for a loading increases as the factor number increases. This is chiefly necessary because, as noted earlier, the proportion of unique variance which creeps into all factors is so great in later factors as to swamp the common variance.

With sample sizes of 77 (boys) and 84 (girls) the significance levels of Pearson's 'r' correlation coefficients are 0.22 and 0.23 at 5% and 0.29 and 0.31 at 1% for boys and girls respectively. Considering, however, that the unique variance in the present analysis has been extracted before computing the factor loadings, the level at which a loading becomes significant is likely to be reduced accordingly. The 1% level is therefore likely to be unnecessarily stringent, and a 'middle' arbitrary figure of 0.25 is therefore taken as a basic level at which loadings are likely to become significant in the present study.

A larger sample size would give more power to the results, but with the previous evidence suggesting differences in the creative interests and activities of boys and girls, and the inclusion of personality and attitude variables, a factor analysis is undertaken for both sexes separately. The results are considered in the next section.

11.2 RESULTS AND DISCUSSION

Results of the factor analysis and the associated correlation matrix, for both sexes, are reproduced in full in the Appendix. For ease of discussion the factor matrices are also presented in Tables 40 and 41, giving only those loadings which are of significance, near-significance or particular psychological interest. The tables are divided into five main areas of convergent thinking, attitudes, divergent thinking, creative criteria and personality variables, reading from top to bottom.

Before the extraction of unique variance, the analysis yielded eight factors for girls and nine for boys accounting for 72.0% and 74.5% of the total variance respectively. These percentages were reduced to 62.7% and 64.1% respectively when only common variance was considered, illustrating a considerable increase in the 'purity' of the factors from the 'hybrid' factors arrived at by Principal-Components analysis.

Factors from the Boys and Girls analysis will be identified by the letters 'B' and 'G' respectively and will be considered jointly whenever their identity is similar.

Girls (n = 84)

	FACTORS 1	2	3	4	5	6	7	8	h²
I.Q.1	85	25							82
I.Q.2	85								80
English	85	(21)	(17)						82
Maths	87	(17)							82
Attitude to School	36		30		42	37			57
Importance of Doing Well	47				27	39		-28	55
Conforming					76				59
Relationship with Teacher	39		(17)	(22)	48	25			54
Lack of Anxiety								59	37
Academic Self-Image	47				(23)	54			60
Self-Concept			(18)			72			60
Circles Flu.		86							77
Circles Flex.		82							73
Circles Orig.	(16)	67						(18)	55
Uses Flu.		75	(23)					-23	69
Uses Orig.	(17)	67	(19)					(-17)	56
Consequences Flu.		61	(23)			27		(-19)	56
Consequences Orig.	(23)	54							41
Board Game	34	(22)			26	(-21)	27	26	43
Interests (overall)			74			(18)	57		94
Imaginative Interests			87						84
Logical/analytical Interests		(17)	(24)				83		78
Creative Behaviour Teacher Ratings	69	(16)	(17)	-37					70
Leisure Interests Torrance		38	46			30	(19)		54
Creative Motivation Golann	28	(20)	46	(-21)	(20)			(19)	46
Extraversion	(19)	(17)		(-23)	(-23)	47		32	51
Neuroticism		27		(22)	(-27)			-53	49
Lie Scale	(-23)			(24)	52				40
Shy				93					90
Lively				-73	(-19)				62
Pleasurability	47			38	24				50
Percentage of Common Variance	25.0	21.4	11.5	10.5	9.6	9.1	6.5	6.2	
Percentage of Total Variance	15.7	13.4	7.2	6.6	6.0	5.7	4.1	3.9	

Common Variance extracted = 62.7%

(decimal points are omitted in the table)

Table 40 Principal-Factor Analysis with Varimax Rotation

Boys (n = 77)

	FACTORS 1	2	3	4	5	6	7	8	9	h^2
I.Q.1	86						(16)			82
I.Q.2	79			(21)			(23)			77
English	87	(23)								83
Maths	85		(18)							77
Attitude to School		(20)	43	60	(16)					63
Importance of Doing Well	32			65						57
Conforming				76						60
Relationship with Teacher		(-18)	42	48				(21)		52
Lack of Anxiety	27		65				26			60
Academic Self-Image	53		40	(18)				29		61
Self-Concept			70					(18)	(19)	64
Circles Flu.		90					(19)			87
Circles Flex.		91								87
Circles Orig.	26	67				(22)				60
Uses Flu.		44					71			73
Uses Orig.	32	26	(20)				40	-34	(20)	53
Consequences Flu.	25	72								63
Consequences Orig.	43	49								47
Board Game	25	(17)		(16)		32	52			51
Interests (overall)					93			30		99
Imaginative Interests					85					80
Logical/analytical Interests	28				31		25	54		56
Creative Behaviour Teacher Ratings	69					48			(24)	81
Leisure Interests Torrance		27		-27		(21)	(23)			29
Creative Motivation Golann									57	40
Extraversion	(16)		68	-26		(18)	(-19)		(22)	70
Neuroticism	(-16)		-31			(21)		-31	(23)	33
Lie Scale	(-17)			(17)				43		30
Shy		(-19)			-25	-76				71
Lively	25	(20)				79				77
Pleasurability	33			33				30	57	67
Percentage of Common Variance	22.4	17.9	11.2	10.5	9.7	9.4	7.2	5.8	5.0	
Percentage of Total Variance	15.0	11.5	7.2	6.7	6.2	6.0	4.6	3.7	3.2	

Common Variance extracted = 64.1%

(decimal points are omitted in the table)

Table 41 Principal-Factor Analysis with Varimax Rotation

11.21 Identification of the Factors

An initial inspection of the factor loadings shows that most factors appear broadly in parallel, in both sexes, though there is also some variation in the loadings. The general identity of the factors can be summarised as shown in Table 42, though the underlying interpretations may be different and will be discussed shortly.

Girls		Boys	
Factor		Factor	
1	General convergent ability	1	General convergent ability
2	Divergent thinking	2	Divergent thinking
3	Creative Interests (Imaginative)	5	Creative Interests (Imaginative)
4	Personality Shy/Lively (bipolar)	6	Personality Shy/Lively (bipolar)
5	Attitudes to School	4	Attitudes to School
6	Self-Concept, Extraversion	3	Self-Concept, Extraversion, Lack of Anxiety
8	Creative Interests (logical)	8	Creative Interests (logical)
8	Lack of Anxiety/ Neuroticism (bipolar)		
		7	Divergent thinking (Uses Test) and Board Game
		9	Creative Motivation and Pleasurability

Table 42 Initial Inspection
Major Loadings or General Description of Factors

Each of the main groups of tests, convergent, divergent, attitudes, personality and creative activities is represented by one or more factors, and the main concern of this study is with the divergent factors and the relationship between divergent thinking and the other main groups of tests. These will be the main issues in the subsequent discussion, though interesting relationships between the other abilities will be noted in passing.

11.22 Discussion

11.221 Convergent Thinking and Divergent Thinking

For both boys and girls the first factor, provisionally labelled 'general convergent ability' is identified by its very high loadings for intelligence, English and Mathematics, all in the region of 0.85. The end of year assessments for English and Mathematics appear as clearly as those of intelligence and the factor fits Vernon's (1961) description of a 'g + v:ed' ability including both 'g' and a general verbal/ numerical/educational ability. In terms of the present study the nature of the school assessments and the tests of intelligence put the factor into Guilford's category of convergent thinking and for each sex it shows only a slight degree of overlap with factors 2G and 2B respectively which can be identified as divergent ability. The latter are clearly recognised with substantial loadings on all seven divergent measures as high as 0.86 for girls and 0.91 for boys.

Together the first two factors account for a large proportion of the common variance, though only 11 out of 31 of the variables are specifically designated as convergent or divergent ability. For girls 46.4%, and for boys 40.3% of the common variance is associated with these factors, the major proportion loading on the convergent ability factor along with the four convergent tests. Apart from Circles (flexibility) for boys the divergent tests all have some positive loadings on the convergent factor, particularly for the originality scores. For girls the latter are of some interest as a group (0.16, 0.17 and 0.23), though they do not individually load significantly. For boys the same scores load 0.26, 0.32 and 0.43 and indicate that Guilford's dimension of divergent production of transformations (see Chapter 3), for both verbal and figural content, shares some common variance with convergent thinking abilities. Fluency scores are not in evidence in the same way, and the originality loadings cannot, therefore be explained simply in terms of verbal fluency. The most likely explanation is that unusual responses to divergent tests involve a process of evaluation as to their appropriateness which utilizes a certain degree of reasoning ability.

The divergent factor is most strongly associated with the figural divergent thinking abilities assessed by the Circles Test, and it is clear that the suggestion made in the last chapter that the flexibility score is closely related to fluency, is borne out by these results. Although there are minor fluctuations it shows little separation from fluency and with almost identical amounts of variance accounted for, and no evidence of different validity associations in the last chapter, there are no grounds for recommending its retention.

In view of the theoretical basis for such an ability argued for in Chapter 3 this result is disappointing, but as pointed out earlier, the scoring procedure for originality automatically demands a high degree of flexibility, and it appears that the latter is more distinct from a fluency ability than the simple change of category involved in the flexibility score itself.

The verbal divergent tests load heavily on the divergent factor but they also, particularly for boys, show some substantial loadings on other factors. These will be discussed when these factors are identified in terms of other variables, but Factor 7B is in fact most clearly recognised by its high loading 0.71 for Uses (fluency). This will also be returned to again later but at present it indicates that divergent thinking is not an unitary ability. While the divergent tests appear more homogeneous for girls, divergent abilities for both sexes load on other factors including that of convergent ability.

In turn the four convergent tests also show positive loadings on the divergent factor, including three loadings of 0.17, 0.21 and 0.25 for girls, and 0.23 for boys. The English assessments are responsible for the loadings of 0.21 for girls and 0.23 for boys and though only on the borderline of significance suggest that teachers do recognise some divergent ability in their assessments of English performance. This provides a slender amount of 'academic validity' similar to that found by Bennett (1973) in relating divergent abilities to conventional English attainment. The conventional English assessments in the present study were of a fairly traditional nature and it is possible that if the attainment was geared to more imaginative assessment, divergent abilities might play a greater part.

In terms of the present results however the most realistic conclusion is that divergent abilities play little part in conventional academic attainment in English and Mathematics, and the latter bears out an earlier conclusion of the writer (Richards and Bolton, 1971).

The relative independence of factors 1 and 2 for both boys and girls and their clear interpretation in terms of convergent and divergent ability allays some of the doubts expressed by Butcher (1972). There is evidence in the present study that a "general factor of divergent thinking" *is* readily distinguishable from the corresponding factor of convergent thinking.

The results can be interpreted, to some extent, in support of Wallach and Kogan's (1966) conclusion that they had succeeded in defining a dimension of individual differences which showed both generality and pervasiveness and at the same time independence of traditional measures of general intelligence. While Wallach and Kogan claimed that their dimensions were quite distinct however the present results indicate that there is also some degree of overlap between divergent thinking and intelligence, and that divergent thinking is also not as unidimensional as they suggest.

Similar conclusions have been reached with data from high ability samples (e.g. Ward, 1967; Cropley, 1968; Starr and Nicholl, 1975), and from unrestricted groups by the writer (Richards, 1970) and Hargreaves and Bolton (1972), but not by Hasan and Butcher (1966). The present finding is slightly more pronounced than in the earlier study of the writer, though a similar interpretation is possible, namely that divergent thinking ability appears as a dimension of intellectual ability, which, while not independent of intelligence, exists as a "consistent complementary activity".

On the earlier occasion the tests were largely cognitive, but the majority of tests in the present study are of a non-cognitive nature and in view of the suggestions by a number of researchers that attitudinal and personality variables are likely to play a large part in divergent performance the relatively distinct nature of the divergent factor is more remarkable.

Some relationship with the non-cognitive variables will be noted shortly when these variables are considered more specifically but there is no suggestion in the present results that divergent thinking is largely a function of attitude or personality. Rather it has been seen to exist largely as a cognitive variable both in terms of its long-term stability and with the present evidence for its construct validity. It is not, as noted above, an entirely unidimensional variable though there is sufficient common variance to suggest the presence of a *cohesive dimension of divergent thinking abilities*.

The predominance of the figural abilities on the divergent factor and the separate loadings noted for the verbal tests on other factors indicate that it would be possible in a modified battery to 'blow up' these aspects in order to emphasise Guilford's multivariate view of the intellect (Guilford, 1967, 1976). While some researchers, e.g. Cattell (1971) maintain that this would be to distort a natural sampling picture of human abilities, it might serve to distinguish some of the main components in 'general divergent ability' in a way which would make prediction easier.

For multiple correlation purposes one would not wish all divergent tests to load on a single factor as this would only duplicate correlations with other variables, and as suggested in the last chapter, further attention might be given to complementary rather than parallel tests of divergent thinking.

Unless one wishes to manufacture a narrow, unitary concept of divergent thinking, it is not true, as Hargreaves and Bolton (1972) suggest, that it is obviously preferable "in selecting divergent creativity tests, ... to select tests which load most highly on a general 'creativity' factor, and which have minimal loadings on others". Starr and Nicholl (1975), for example, demonstrate a considerable factorial separation between verbal and figural tests in high ability subjects, and the present results indicate that different tests can yield different relationships with creative criteria.

At present the distinction between the various divergent abilities is not great, but although there is some basis in the divergent factor for Richards R.L.'s (1976) conclusion that divergent test intercorrelations argue for a "relatively simple description of the creative thinking domain, supporting the Wallach-Kogan position of a unitary and independent dimension", it is not a valid interpretation of the total picture, and ignores the small amount of factorial separation of the divergent abilities both within tests and between the tests themselves. Although there is a stronger grouping of loadings within each test, there is also some separation of these abilities and some evidence that scores, for originality in particular, are related across the test boundaries.

There are slight differences between boys and girls in these respects, but the interpretations of factors 1 and 2 are very similar for both sexes and suggest that, when cognitive tests are being analysed, an overall factor analysis would be appropriate. This reinforces similar conclusions by Cropley and Maslany (1969), Hargreaves and Bolton (1972) and Bennett (1973).

In general the evidence of the convergent and divergent factors provides construct validity of the discriminant/convergent variety, but also suggests that while there is an integrated range of abilities represented by the divergent tests which is relatively distinct from general convergent ability, it is not an entirely unitary concept. This finding fits in well with the theoretical interpretation of creativity in Chapter 2 as a broad rather than an unidimensional trait, and also with the concurrent validity findings in the last chapter.

11.222 Attitudes and Divergent Thinking

For both sexes attitudes load quite heavily on three main factors, on the general ability factor, and on two factors defined by various attitude scores themselves. The latter are defined in a very similar fashion for each sex.

Factors 5G and 4B accounting for 9.6% and 10.5% of the common variance for girls and boys respectively both have their major loading for 'conforming', 0.76 in each case, reflecting children's positive response to good behaviour in the classroom as opposed to 'fooling about' and 'liking noisy children'. Children with conforming attitudes also tend to have good attitudes to school and school work, to be introverted rather than extraverted, to see themselves as trying hard to do well, and as having a good relationship with the teacher.

This is substantiated by the teacher's expression of pleasure to them in class, loading 0.34 for girls and 0.33 for boys on the respective factors. The loading from the Lie Scale of the Junior Eysenck Personality Inventory, of 0.52 for girls and a more moderate 0.17 for boys gives some support to the interpretation of the Lie Scale as an index of the children's attempt to present themselves in a good light, and as a measure of "conformity to social pressures" (Eysenck, S., *et al*, 1971).

In view of the unconventional attitudes and aspirations associated with the divergent pupil, by Getzels and Jackson (1962), and Hudson (1966), the loadings for divergent and convergent abilities on the above factors (5G and 4B), could have been expected to be negative and positive respectively. None of the loadings are significant however, though for boys, there is a slight trend in this direction. Non-conforming attitudes are also only very slightly related to divergent thinking on the main D.T. factor for each sex, and over the whole range of ability in this study, it has to be concluded that non-conforming attitudes are not a positive influence on divergent ability. Vernon (1972) reported a similar finding and even an indication of conforming rather than non-conforming attitudes amongst divergent boys. It is possible that such an influence if it exists, is only apparent amongst groups high in general ability.

Factors 6G and 3B form the other pair of attitude factors accounting for 9.1% and 11.2% of the common variance for girls and boys respectively. Both are well defined by a high loading for self-concept, academic self-image and extraversion, and reflect an accompanying good attitude to school and perceived relationship with the teacher. In contrast to the other factor, the confident, outgoing characteristics associated with this hypothetical trait might be thought to be associated with high divergent performance. Only an isolated loading of 0.27 for Consequences (fluency) for girls suggests that this might be so, and for boys there is quite a substantial negative loading of -0.20 for Uses (originality). The overall pattern has to be interpreted as showing no appreciable relationship with divergent ability.

An interesting difference between boys' and girls' attitudes in the classroom is shown by a high loading of 0.65 indicating 'lack of anxiety' on factor 3B, but no corresponding loading on 6G. Lack of anxiety for girls comprises instead, the major loading on factor 8 for girls. In both cases lack of anxiety on the Barker-Lunn scale is also reflected by a negative loading for neuroticism on the Junior Eysenck Personality Inventory, indicating relaxed and carefree rather than anxious behaviour.

For boys, lack of anxiety is associated with high self-concept, extraversion and good attitudes to school, but for girls it shows little relation to other attitudes about school work. The only associated loading with this easy-going behaviour in girls is a negative value of -0.28, indicating a lack of acknowledgement of the importance of working hard at school work. Looked at from the point of view of low scorers it appears that anxiety is more related to boys' attitudes than girls' attitudes at this age, anxious boys appearing to be intraverted with poor attitudes to school. Which, if any, of these are causal factors is a matter for further study, but the results may be relevant to the finding of Eysenck and

Cookson (1969), that the commonly reported link between neutoticism and poor attainment at primary school was particularly marked for boys.

In addition to these specific attitude factors, attitudes load quite extensively on factor 1 for both boys and girls, though by contrast there are no significant attitudinal variables loading on the divergent factor.

Children's attitudes and convergent ability are related by their positive expression of the 'importance of doing well' (e.g. 'I work and try very hard in school') and a high level of academic self-image (e.g. 'I'm good at sums'). It appears that the successful children by school standards both work hard and achieve self-recognition for their performance. They also tend to be seen by the teacher as being a pleasure to have in class, loading 0.47 for girls and 0.33 for boys on factor 1. While the teacher's interest is recognised by the girls ('Perceived relationship with the Teacher') it is not recognised by boys. The more able girls also tend to have better attitudes to school though again this is not true of boys. The relationships with the convergent variables are generally similar to those of Barker Lunn (1970). Although her analysis was conducted for a combined population, a correlation between sex and the variables in question also suggested that like here, the relationship between 'attitude to school' and academic work is stronger for girls than boys.

Though not significant, there is an indication on the divergent factor for boys that a good attitude to school is associated with better divergent performance (0.20), and a further small but interesting loading of -0.18 suggests that the higher boys divergent performance the less likely they are to feel that 'Teacher is interested in me'. This is borne out by the corresponding correlation coefficients, all seven coefficients between the divergent tests and boys' perceived relationship with the teacher being quite substantially negative, two, with the Uses Test, being significant at 5%. Even if the teacher is in fact interested, there is a suggestion that divergent boys are less likely to perceive it and that there might consequently be need for more encouragement for divergent type responses.

Teachers' ratings of pleasurability are also closely related to convergent rather than divergent ability, and the loadings on factors 1 and 2 for both boys and girls could be interpreted as giving support to the findings of several writers, (Getzels and Jackson, 1962; Hasan and Butcher, 1966; Cropley, 1967), that teachers find divergent children less desirable as pupils than convergent children. As defined by these researchers and earlier in this study by the writer, the term 'divergers' however excludes the 'all rounders' category and automatically means that divergers do not appear in the top 20% or 30% on I.Q. The evidence of the present study suggests that it is the absence of high intelligence rather than the possession of high divergence that is an important factor in teachers' ratings of pleasurability.

As Biggs *et al* (1971) points out however, even if teachersers see high divergers as undesirable pupils, the more important issue is whether this is because they cannot perceive children with high divergent ability realistically, and consequently underestimate their ability and misinterpret their classroom behaviour. As discussed earlier, with regard to teacher ratings in Chapter 4, Biggs found that teachers did perceive children with divergent ability veridically, evaluating their abilities and behaviour as positively and realistically as those of the convergers.

Though the ratings for 'mechanicalism' were not very clear-cut, Biggs concluded from the 'conceptualism' rating that teachers took both convergent and divergent ability into account, and he also noted that 'good' behaviour did not appear to be especially related to convergence, nor 'nuisance' behaviour to divergence.

The loadings for teacher ratings of 'pleasurability' and 'creative behaviour' on factors 1 and 2 suggest an interpretation mid-way between the positive findings of Biggs and the contrary suggestion that divergent ability is discriminated against. There is no indication in the present findings that divergent thinking ability is a handicap to either rating, but it is also clear that unlike Biggs' conclusion divergent thinking is not seen as a positive advantage, and pleasant relationships and creative behaviour are clearly associated with general convergent ability.

Considering the predominantly traditional methods of the teachers however the modest positive nature of three out of the four ratings in question leans slightly towards Biggs' findings, and with teachers less formally inclined the recognition of divergent ability might be more marked.

11.223 Creative Activities and Divergent Thinking

It was noted in the last chapter that although divergent tests had a significant degree of concurrent validity in relation to creative activities there was a good deal of variation in the pattern of relationships. This is borne out by the factor analysis which underlines the multidimensional nature of 'creativity'. Although all the creative criteria has positive loadings on the divergent factor for both boys and girls these vary in size, and most of the creative activity variance is spread over other factors, in some cases with accompanying loadings from the divergent tests.

For girls Torrance's leisure interests has a loading of 0.38 and four out of the remaining six 'creative' loadings, 0.16, 0.17, 0.20 and 0.22 are substantially positive, the latter two being of borderline significance. Creative activities also load on the convergent factor, but apart from Teacher Ratings these are overall no more marked than on the divergent factor. Factor 3G is a clear creative activities factor but like 7G it is somewhat 'artificially' produced by the inclusion of both part scores and total scores for the Barker Lunn Interests questionnaire.

Factor 3G links together the self-report inventories of creative interests and motivation with several moderate loadings on the verbal divergent tests (0.23, 0.19, 0.23, 0.09), which though not individually significant comprise a group of likely psychological significance. It must be remembered that with uncorrelated factors this common variance exists over and above that already extracted by the convergent and divergent factors and, with the low loading of 0.17 for English ability suggests an underlying imaginative ability of a verbal nature, some of which is recognised in the small residual loading (0.17) on the Teacher Ratings for Creative Behaviour.

Factor 7G emphasises some of the independence of logical interests from imaginative interests, but exists largely because of its overlap with the total interests score. It has a significant loading from the Board Game (0.27), indicating that logical/analytic interests increased girls' chance of success at the Board Game, but has no other significant loadings.

Performance on the Board Game, for girls in particular, is well spread between the factors and only a relatively small proportion of its variance is extracted. This exemplifies Guilford's (1976) conclusion that not even a collection of all the divergent production abilities, will completely account for creative aptitude. While Board Game performance has moderate positive relationships with girls' convergent and divergent performance it also has loadings on several other factors which have no significant divergent loadings, and in the case of factors 6G and 8G has some negative but non-significant associations with certain aspects of divergent performance. In the latter cases and on factor 5G, success at the Board Game is more closely related to attitudes and personality variables than divergent thinking but in no case is the relationship very marked. The overall pattern however gives some limited support to the views of several writers (e.g. Butcher, 1972; Nuttall, 1972) that attitudinal and personality factors are likely to be of importance in practical creative activity.

For boys the loadings for creative activities are similarly diffuse, but with some different patterns of relationships. On factors 1 and 2 similar interpretations can be made as those for girls, with a high loading for Teacher Ratings on factor 1 and otherwise small but overall positive relationships, but the remaining picture is rather different. Factor 5B like 3G is an imaginative interests factor but unlike 3G has no relationship with divergent abilities. Factor 8B contains the same 'artificial' pairing of logical interests with the total interests score as on 7G, but it is also associated with stable and conforming behaviour, and a negative relationship with Uses (originality). While logical interests have some negative association with divergence on this factor, creative motivation on factor 9B is positively associated with the same divergent score. Although the variance extracted by these factors is small their different relationships with divergent thinking indicates the multifactorial nature of both creative activities and divergent thinking.

The most significant relationship between the creative criteria and the divergent abilities is shown on the secondary divergent factor for boys, 7B, which is predominantly a verbal divergent factor but has high loadings from the Board Game (0.52) and moderate loadings of 0.25 and 0.23 from logical/analytic interests and Torrance's leisure interests respectively. Considering that boys performance on the Board Game is also related to a limited extent with convergent and divergent abilities as shown by factors 1 and 2, this is the most significant link between a distinct ability involving the fluent and original production of ideas and practical creative performance. It cannot be interpreted in terms of general exuberance or extraversion but is instead weakly related to introversion, a lack of anxiety, logical interests and to a small extent I.Q.

The Board Game called for a combination of unusual uses for objects, an evaluation of their suitability, and the ability to put the ideas into practice. It appears that a combination of such abilities can be recognised in terms of a hypothetical trait underlying divergent production on the Uses Test, a degree of thoughtful introversion, lack of anxiety, reasoning ability, and logical/analytic interests. This accounts for 25% of the variance in the Board Game, 15% of which is provided by the fluency score for Uses. This provides additional support for the concurrent validity established in the last chapter and suggests that divergent tests do more than credit "mere glibness, uninhibited self-expression and deficiency of self-criticism" (Ausubel, 1968).

11.224 Divergent Thinking and Personality Variables

A number of significant personality loadings have already been mentioned, but their influence on divergent thinking performance has been slight. The mainly attitudinal factors, 6G and 3B, with high loadings for extraversion (E), and for boys, an emotionally stable loading for neuroticism (N), had to be interpreted as showing no appreciable relationship with divergent thinking in spite of their confident, outgoing characteristics. The other pair of attitude factors, 5G and 4B, reflected intraversion and, for girls, neurotic behaviour, but again showed no relationship to divergent thinking. While personality variables are therefore related in quite interesting ways to various of the attitude scales, they have so far appeared to exert little influence on divergent performance.

Extraversion has a low loading of 0.17 on the divergent factor for girls and the corresponding correlation coefficients are similarly low but positive, one of the seven coefficients reaching significance at 5% and three others being borderline.

For boys there is a tendency the other way, extraversion loading -0.12 on the divergent factor, 2B, and the accompanying correlations also low and generally negative. Taken together with the above findings for the attitude factors carrying loadings for E, children's level of extraversion is clearly not a deciding factor in determining their divergent performance, though it appears to be slightly more of an advantage for girls.

Though it has a zero loading on the divergent factor for boys, neuroticism is a significant variable for girls' divergent performance, loading 0.27 on 2G. This is contrary to the general finding that neuroticism tends to have a negative effect on performance at primary and lower secondary level (Eysenck, 1972; Entwistle, 1972) though performance is generally assessed in terms of academic attainment. Neuroticism *is* in fact slightly negatively related to convergent performance for boys on factor 1B (loading -0.16) and there is also limited support on both 1B and 1G for the general finding that extraverts seem to have an advantage in general academic work.

The positive advantage given by neuroticism to girls divergent thinking is also indicated on factor 8G. In this case it is related to verbal rather than figural aspects of divergent performance and this is borne out by the correlations between N and DT. All four of the verbal test correlations are positive and significant, two at 5% and two at 1%, while those with the figural tests are also positive but with only one coefficient nearing significance.

It is interesting to speculate that this change from the normal negative relationship of neuroticism to academic performance reveals a consistent difference in performance when the material requires divergent rather than convergent performance. The 'drive' that comes from mild anxiety may be indicative of 'creative energy', though for both boys and girls there is only a very slight indication that success at the Board Game, for example, is enhanced by a degree of neuroticism.

The relationship between divergent performance and neuroticism deserves further investigation in relation to Eysenck's (1972) emphasis of Hull's general learning theory in terms of performance as a product, D x H, of motivation/drive, D, and learning/habit, H. If as Eysenck suggests, learning depends on the strength of existing habits as well as on the degree of drive, then a variation in performance when the stimulus calls for divergent rather than convergent performance might be expected. It is also possible that the finding that neuroticism is an advantage to older students is related to the same aspect, performance at the primary and secondary stage normally being assessed in terms of convergent ability, while at high levels more imaginative and divergent performance may gain credit.

Entwistle (1972) reports on a number of studies that have associated neuroticism with subject choice, and Hudson (1968) while suggesting that neuroticism in convergers and divergers takes different forms, found divergers to be significantly more neurotic.

Hudson's study was with a group of sixth-form boys however, and his findings were not supported by Smithers and Child (1974) with a sample of university students of both sexes. The present findings extend those of Hudson to a sample of 11-year-old girls but do not confirm his conclusion for boys.

Teacher Ratings of the degree to which their children appeared to be 'Shy' or 'Lively", were included in the belief that they might help to establish a personality dimension, and make it easier to interpret subsequent relationships between such a dimension and divergent thinking. The result however is largely confined to a bipolar personality factor with 'Shy' at one extreme and 'Lively' at the other. Without details of the emotional behaviour and activities that the writer felt would be associated with these descriptions it was perhaps unreasonable to have expected teachers to do other than see them as almost mutually exclusive.

For both boys and girls the corresponding factors 4G and 6B carry loadings of only borderline significance from the Extraversion and Neuroticism scores on the JCP I. As one might expect loadings for extraversion are in the direction of 'lively' for both sexes, and shy as opposed to lively girls appear as slightly neurotic. Somewhat surprisingly the opposite trend appears for boys. The terms 'lively' and 'shy' may however convey somewhat different meanings to teachers depending on which sex the terms are used to describe.

In spite of rating shy girls as being more of a pleasure to have in class, teachers still rate the lively girl highly in terms of creative characteristics, giving some further weight to the belief that teachers ratings of performance are not dictated by personal preferences. There is no association between the 'shy' or 'lively' characteristics and pleasurability for boys, though lively boys are also seen by teachers as potentially more creative.

This popular association of liveliness with creative activity is not given much support by the present study, none of the creative criteria being associated with liveliness in girls. For boys however success at the Board Game has a significant loading of 0.32 towards the 'lively' pole of factor B6, reflecting correlations of 0.33 and -0.30 with 'lively' and 'shy' ratings respectively. There is also an accompanying loading of 0.22 from Circles (originality). Whether a lively/shy rating will predict success on any other practical criteria is a matter worth investigating further, and any similar relationships for girls may be more likely to appear if the criterion were designed with more of a feminine flavour. Apart from the one borderline loading of 0.22 on B6 however there are no links between the shy/lively dimension and divergent thinking.

It is noticeable that apart from the heavy loading for extraversion on the attitude factor 3 for boys, both extraversion and neuroticism are characterised by several moderate loadings rather than by any specific factor, and that the Board Game, especially for girls, is rather similar. The amount of variance extracted for these variables is also relatively small and it is clear that both practical performance on the Board Game - the criterion nearest to a 'real-life' situation - and measures of extraversion and neutoticism, are far from being related to other variables in any close or clear-cut way.

Their inter-relations as measured by the correlation coefficients are low and in three cases out of four negative, but in the factor analysis some positive groupings are also in evidence. This suggests that the relationship may not be linear or that the effects of E and N depend on their grouping with other variables. Lynn and Gordon (1961) suggested that a curvilinear relationship to attainment was a feature of neuroticism, though there is no suggestion of such a relationship in the results of Eysenck and Cookson (1969), and Entwistle and Cunningham (1968) demonstrated linearity of regression among their findings. Grouping of contrasting personality types in both studies however revealed results differing from the overall picture and between the two studies. As Eysenck (1972) notes, knowledge of the relations between academic attainment and personality is far from certain, though he suggests that the evidence indicates that there are likely to be important relationships in this area. It is likely that relationships between divergent thinking and personality may also be revealed in a study of selected groups contrasting on other variables.

Over the whole range of ability represented in the present study however it has to be concluded that personality variables do not play a major part in divergent thinking performance. Though there is some significant association between neuroticism and divergent thinking for girls. Eysenck's (1967) suggestion that divergent thinking is largely a function of children's extraversion is firmly rejected.

11.3 SUMMARY AND CONCLUSIONS

The main conclusions in this chapter can be summarised as follows:

1. There exists a cohesive dimension of divergent thinking abilities which is relatively distinct from general convergent ability.

2. While divergent thinking abilities have a great deal of common variance they do not form an entirely unidimensional variable, and further attention might be given to complementary rather than parallel tests of divergent thinking. This is particularly true of figural and verbal aspects of divergent ability, and different divergent thinking abilities may be found to be relevant to different creative criteria, as suggested in the last chapter.

3. The flexibility score for the Circles Test adds so little information to the fluency score for the same test that its retention in future studies cannot be recommended. Fluency and originality scores, while strongly related within tests, are by no means identical and though a fluency score is likely to indicate a considerable proportion of children's divergent potential, the originality score is recommended for any detailed study.

4. Attitudes fall into two main categories, positive attitudes to school going with conformity, effort to do well at school work and slight introversion on the one hand; and self-concept, extraversion and academic self-image on the other. Although this dichotomy appeared to reflect characteristics of two different types of children within each sex, neither had more than minor relationships with divergent performance.

5. There was no indication that children with high divergent thinking ability are discriminated against when teachers assess their desirability as pupils. Teachers preferences were on the other hand positively related to children's convergent thinking abilities, conformity and good attitudes to school.

6. The factor analysis underlined the multi-dimensional nature of creative activities, and the fact that such activity is unlikely to be accounted for by any particular group of variables divergent or otherwise. Divergent thinking abilities nevertheless accounted for a significant amount of creative variance, though this was due to some extent to the corresponding multi-dimensional nature of the divergent abilities rather than to the two main divergent factors. Factor 2 for girls, however, indicated the presence of an imaginative ability linked to verbal tests of divergent thinking and creative interests, and factor 7 for boys provided a strong link between the divergent production of fluent and original ideas and practical performance on the Board Game.

7. Over the whole range of ability represented in this study, personality variables do not exert a major influence on divergent performance. In particular Eysenck's (1967) suggestion that divergent thinking is largely a function of children's extraversion is firmly rejected. Hudson's (1968) finding that neuroticism was positively related to divergent thinking performance in boys was not confirmed, but a similar finding was established for girls.

The latter is in the opposite direction to the commonly reported influence of neuroticism on convergent performance at this age, but in the same direction as that found for older students. It is suggested that the effect of personality may depend on the type of material dealt with, and that some of the age variation may therefore be due to a change of emphasis in the type of performance required. This deserves investigation within selected groups with different combinations of personality.

8. Teachers tend to rate lively children as more creative, and in the case of boys this was borne out by the boys' performance on the Board Game. Teacher ratings in this respect deserve further investigation.

9. Considering the similarity of the factor patterns for each sex an overall analysis would have given rise to very similar conclusions regarding the relationships between convergent and divergent thinking. A number of the relations between these variables, creative activity, personality variables and attitude however would have been obscured in a combined group. As noted in the last chapter, when creative activities and interests are concerned there is still a good deal of truth in the assumption that there are activities 'typical' of either boys or girls.

These findings provide a picture of the construct validity of divergent thinking tests in terms of a dimension of cognitive abilities, relatively independent of intelligence and not, as some writers have suggested as a function of personality and attitudes. While, as Hudson (1968) observed, some lack of convergent and discriminant validity has led to arguments that divergent thinking cannot be a coherent human trait at all, the present results indicate otherwise. Though not as unidimensional a trait as suggested by Wallach and Kogan (1966) or Richards R.L. (1976) divergent thinking is clearly an aspect of intellectual ability that deserves attention in its own right.

Taken in conjunction with the earlier evidence regarding reliability and long-term stability, the present results give support to the realistic nature of a concept of divergent thinking as one component in a broad conception of the intellect.

That divergent thinking is not highly correlated with personality or attitudinal variables does not mean that the latter do not exert some influence or are unimportant. On the contrary the influence of personality on divergent thinking appears to be of a similar order to its influence on academic attainment and I.Q., and though small in an unrestricted population no one would deny the importance of the latter. With the evidence of this chapter supporting a construct of divergent thinking as a cognitive dimension, it is likely that like other cognitive variables its effect will be mediated in numerous ways by the variations in children's attitudes and personality. Further investigations of divergent thinking, creative performance and personality variables will be needed to discover particular groupings that may be most advantageous for different types of learning and creative activity.

It would not be realistic to suggest that divergent thinking is a major influence on creative activity, but the present evidence gives some basic support to teachers who regard the fostering of divergent thinking abilities as a realistic and worthwhile aim.

CHAPTER TWELVE

OVERVIEW AND CONCLUSION

In many ways it is easy and enjoyable to be carried along on the bandwaggon of the latest psychological fad or educational fashion, and the subject of creativity as it emerged during the early 1960's was something of both. Since reading 'Creativity and Intelligence' (Getzels and Jackson, 1962), 'Modes of Thinking in Young Children' (Wallach and Kogan, 1966) and 'Contrary Imaginations' (Hudson, 1966) nearly ten years ago, the writer has enthused about the virtues of divergent thinking and the open-ended, flexible and child-centred methods of teaching that, as indicated at the beginning of this study, have been associated with the concept of creativity.

Having had a long run however there are now signs that the bandwaggon is slowing down, having, some critics suggest, run markedly out of control in the meanwhile (e.g. Pole, 1969). Enthusiasm is certainly no longer sufficient to sustain developments, and criticisms, whether of classroom practice or associated psychological principles, have to be appraised honestly. Differences in the allegiance, tradition and temperament of researchers however make completely honest appraisal difficult, and the search for meaning in data, as Hudson (1972) observes, is likely to involve all of us in distortion to a greater or lesser degree. It is hoped that in spite of his wide use of divergent tests the writer's attempt at an honest appraisal has not been so self-deceptive as to promote one particular view of reality at the expense of all others.

The particular criticisms and doubts which provided the focus for the present research centre on the reliability and

and validity of tests of divergent thinking and their labelling as tests of 'creativity'.

In many ways, as outlined in Chapter 2, the concept of creativity has been closely related in educational and psychological practice, Torrance (1963), for example, suggesting that in order to develop his creative abilities a child should learn by "questioning inquiry, searching, manipulating, experimenting, even by aimless play; in short, by always trying to get at the truth", and Rogers (1954, 1970) defining creativity in terms of "emergence in action of a novel relational product, growing out of the uniqueness of the individual on the one hand, and the materials, events, people or circumstances of his life on the other".

In terms of his definition Rogers sees creativity as a personal construct, different only in degree between a child inventing a new game and Einstein formulating the theory of creativity, and this personal level creativity as something possessed by all of us to some degree, is also fundamental to Guilford's (1950) hypotheses regarding creative ability and to Bruner's emphasis on discovery learning (1960, 1961). It also underlies much of the change of emphasis in teaching methods and curricula that has taken place over the last twenty years, but both the psychological principles and classroom practice have been criticised as being trivial, confusing, meaningless and even anti-educational interpretations of the term 'creativity' (e.g. White, 1968; Shouksmith, 1970).

As noted in Chapter 2, to accept the desirability of teaching for creativity is a long way from knowing how best to recognise and foster creative abilities, and it was suggested that an emphasis on the types of thinking that appear to be involved in a broad conception of creativity would be more profitable than an attempt to narrow the definition of creativity, based for example on a product judged by society to be original and valuable.

Adopting in turn a view of creative thinking as a multifaceted activity it was suggested that thinking abilities that have been variously labelled as 'productive', 'adventurous', 'lateral', or 'divergent', and also 'convergent' thinking, could all claim relevance to the creative process, though in the first instance the degree to which they appeared to be worth investigation in this context would depend on establishing a theoretical basis for their construct validity (Cattell, 1964).

In line with a broad multi-dimensional view of creativity it was argued (2.2) that the wide use and identification of divergent thinking tests as tests of 'creativity' (e.g. Wallach and Kogan, 1966; Torrance, 1974; Brown, 1976) was unwarranted, and that it might well be anti-educational if divergent thinking was to be regarded as synonymous with creative thinking to the exclusion of other dimensions in Guilford's Structure-of-Intellect model (Guilford, 1956, 1967).

It was maintained however that this did not mean that the distinctive feature of divergent thinking tests, the ability of subjects to produce a large number of varied and unusual responses to some stimulus situation, was not relevant to creative activity. On the contrary it was argued in Chapter 3 that there is a great deal of theoretical and anecdotal evidence to suggest that such an ability is an ingredient in the thinking processes which have been described as characteristic of those whose creativity is beyond doubt, and which others maintain are essential for cognitive excellence.

Reviewing the theoretical basis for this claim to construct validity the writer concluded that theoretical approaches to creative thinking lent support to the earlier definition of creative thinking as a broad rather than an unitary trait, and also that divergent thinking, with an emphasis on fluent, imaginative, and flexible thinking is likely to be an important element in the process of creativity in its broad sense. While providing grounds for believing that creative activity is unlikely to be encapsuled within the conventional view of intelligence and I.Q. tests, it also emphasised that the latter have an important role to play and that, as Guilford (1976) has been at pains to point out, creative ability is not tied exclusively to his divergent category.

It was also noted that creative aptitude is likely to be related to attitudinal and temperamental variables and that such variables may also affect divergent thinking performance, Eysenck (1967) in particular suggesting that the latter may be largely a function of children's extraversion. Within the more cognitive approaches to creative thinking, it was suggested that the Gestalt approach to problem solving offered a great deal of relevant information, problems, as Wertheimer (1949, 1961) points out not being confined to the schoolroom. The emphasis on a type of thinking which needed to get away from a one-sided inflexible view, and to appreciate changes in meaning of the constituent ideas in accordance with different relationships to the whole structure, was noted as corresponding to a number of aspects of divergent thinking, and to be of relevance to a practical exercise, the Board Game, to be constructed for the experimental part of the study.

While the theoretical evidence provided encouragement for believing that divergent thinking is of relevance for creative thinking it was emphasised that empirical evidence was needed to justify the hypothesis.

The basic nature of the procedures involved in validating psychological tests and the associated problems where D.T. tests are concerned had been discussed in Chapter 1; and a review of previous practical investigations into the reliability and validity of D.T. tests was made in Chapter 4. Considering the problems involved in scoring divergent thinking tests, the varied nature of reliability information, and the limited reports of long-term stability, and predictive validity, it was decided to attempt an investigation of each of these aspects. The predictive validity was confined to concurrent aspects however, but a practical exploration of concurrent validity was also planned to investigate divergent tests in relation to a range of other variables.

It is appreciated that one could justify a separate study of any one of these aspects, but they are so interrelated that it was felt that a single study could look at each of these aspects to a worthwhile degree and still remain of reasonable proportions. In particular the writer was somewhat sceptical of the reliability and validity of scoring procedures especially when based on standard lists of American responses, and an investigation of methods of scoring should, it was felt, be a necessary part of any investigation of reliability and validity.

This was consequently undertaken as the first part of the practical investigations, and included reviewing and establishing scoring procedures for fluency, flexibility and originality, as well as an assessment of inter-scorer reliability. This section of the investigation was based on the 'main' population of 176 local 11-year-old school children, as were the other sections except for the investigation of long-term stability. The latter involved a follow-up study after an interval of approximately five years, of 173 pupils previously tested in 1969, and of mean age 15 years 10 months when retested in 1974. Details of both samples, the tests used and the experimental procedures are given in Chapters 5 and 8 respectively.

It is hoped that the discussion and clarification of scoring procedures carried out in Chapter 6 will be of use to other researchers scoring divergent tests, and particular attention is drawn to the principle suggested for awarding originality marks, which was considered more valid than methods adopted by some investigators. This maintains that specific responses such as a King's face, a golf ball or Mars, should be recorded in terms of their general grouping as 'faces', 'balls' and 'planets' respectively, and given originality marks according to the frequency of the group rather than that of the individual response - unless the latter possesses some special defining characteristic that identifies it in a different relation to the requirements of the test than other members of the group.

A synopsis of scoring details is given in Section 6.5, and these were subject to investigation for interscorer reliability. Though a number of researchers (e.g. Hudson, 1968; Vernon, 1971) point out the subjectivity involved in scoring divergent tests, most researchers tend to accept the claims of Torrance (1974) that divergent tests can be scored with reasonable consistency. Considering the labour involved in scoring D.T. tests the writer was somewhat suspicious, however, of the 'convenient' assumptions of scorer reliability.

In spite of this concern and comments that "there are still serious problems in deciding when responses are different or unusual" (Vernon, 1971), the results indicated that the writer's doubts were largely unfounded. All seven coefficients of interscorer reliabilty exceeded 0.90, and the three fluency coefficients equalled 0.99.

Variations in the scores awarded to certain individuals however, were noted, and some explanations offered in terms of the grasp of the scoring procedures and, perhaps more importantly, in the different degrees of tolerance and appreciation shown by the scorers for responses that at first sight might appear silly or irrelevant. It was suggested that if information on particular scripts is required they should be marked in conjunction with a second scorer, though in general the overall agreement between the scorers was very encouraging and was considered more than sufficient to allow the next stage of the study to proceed on a well-founded and reliable basis.

The coefficients of test-retest reliability were not as encouraging and were, as expected, less than the coefficients of 0.90 and over, frequently reported for internal consistency methods. The limitations of the latter as applied to D.T. tests were discussed in Chapter 1, and the test-retest coefficients were considered to be more realistic estimates of the reliability of divergent tests.

Though not consistently high however, the size of the coefficients which ranged from 0.50 for Uses (originality) to

0.82 for Circles (fluency), were considered to be sufficient to indicate that divergent thinking tests can play a significant part in making decisions about groups of children, though their use in identifying individuals is more doubtful. It was suggested that the lowest coefficients, for the Uses Test, were possibly affected by the motivational differences between the first and second testing sessions and that they may be an underestimate of the test's reliability. Motivational effects in general were felt to be an important variable as suggested by Torrance (1974).

The level of the coefficients were within the range of test-retest reliabilities reported by other studies, however, and direct comparisons were possible with a number of studies involving the same tests with a similar age group. In general the pattern of correlations, means and standard deviations was similar to that which would be expected of most cognitive variables and the level of reliability not dissimilar to that reported for more established tests.

Having arrived at a measure of their reliability the writer was in a position to make a correction for the attenuation of subsequent correlations with the divergent tests, and this enabled an 'ideal' estimate to be made of the intrinsic relationship between the variables in the validity and long-term stability investigations.

After an interval of nearly five years it was satisfying to locate 150 of the original 173 pupils chosen for the long-term investigation. Data was finally obtained from 139 of these and coefficients of stability for the whole sample and for boys and girls separately were all significant, and similar in magnitude for boys and girls. Although not exceptionally high the coefficients ranged from 0.40 to 0.47 for the individual D.T. abilities and from 0.47 to 0.57 for the combined scores.

These were in close agreement with the only comparable English study found reported, and the present sample was more representative of the whole ability range. Although there was some overlap with general intelligence, partially out the effects of the latter still left very substantially coefficients ranging from 0.35 to 0.47 for the individual abilities and it was concluded that the stability could not be put down to the effects of intelligence. Correction for attenuation indicated that the results were potentially even more significant, and intercorrelations between the D.T. tests in 1969 and again in 1974 showed a very similar pattern and gave further support to the stability of the underlying construct.

A consideration of the levels of performance in 1969 and 1974 showed that the divergent abilities underlying the tests showed reasonable development over time and together with the stability in relative ranking suggested that divergent thinking develops and functions in a similar way to other mental abilities.

At this stage of the experiment it remained to be seen whether any concurrent validity could be demonstrated, but if so, divergent thinking could be regarded as an ability worth further attention not only in psychological investigations but in recognising and developing children's creative potential.

Concurrent validity was investigated in Chapter 10 in relation to three independent measures of children's creative interests, teacher ratings of their creative behaviour, creative motivation, and an experimental task, the 'Board Game' designed to involve flexible and imaginative aspects of children's thinking. Details of the testing instruments and the development of the Board Game, including a preparatory lesson on a simple electric circuit and things that conduct electricity, were given in Chapter 5. Apart from teacher ratings and the Board Game the other criteria were assessed by self-report inventories, while the teacher ratings were given on a structured questionnaire constructed by the writer from Torrance (1967). Each child's performance on the Board Game was assessed individually and the quality of performance assessed as described in Chapter 5, from α-responses which involved insight and a fine grasp of the problem, to β-responses which did not end in any constructive result. Up to about 35 minutes was allowed for each child, and photographs of some typical solutions are given in the Appendix.

In attempting to establish evidence for concurrent validity it was acknowledged that there is likely to be some variation in the amount of weight that can be given to different criteria, and their relative merits were discussed when the results were presented. Special emphasis was placed on the Board Game which added a more realistic dimension than found in most validation studies. The main conclusion from the evidence presented was that divergent thinking abilities have some positive validity in relation to creative behaviour.

It was not intended to establish an unidimensional battery of creative criteria, but all the intercorrelations except one were positive for both boys and girls, and the overall validity index of the divergent test battery in relation to a combined creative criterion was 0.44 for boys and 0.46 for girls, both highly significant ($p < 0.001$). The results gave support to previous studies which had found positive evidence for the concurrent validity of D.T. tests, and the general level of the relationship was very similar to that in the small number of studies in which direct comparison was possible.

A number of variations in the validity of the individual divergent abilities were also observed, and though a considerable degree of validity was common to each ability and to both sexes, some different associations would have been obscured in an exclusive use of combined scores and combined sexes. It was suggested that to gain maximum predictive validity from D.T. tests the activity being predicted should be related to the varying interests of boys and girls, and that different D.T. abilities might also be more closely related to certain activities than others. Verbal but not figural tests for girls, for example were closely related to their imaginative interests. As noted in the reliability investigation the question of motivation is again likely to be of considerable importance.

While statistically significant however it has to be emphasised that the magnitude of the validity relationships is not high enough to suggest that D.T. abilities play more than a small part in predicting creative activity and there are no grounds for believing that divergent thinking is synonymous with creative ability. As both Anastasi (1968) and Cronbach (1970) point out however even tests with validity coefficients as low as 0.20 can make a valuable contribution to a battery of tests especially if they are not highly correlated with other tests in the battery, and it was suggested that it is in this light that the level of concurrent validity established in this study can best be interpreted.

Had the I.Q. scores showed a much greater relationship with the creative criteria than the D.T. scores the proportion of variance contributed by the latter might have been comparatively unimportant. Apart from teacher ratings however, the magnitude of the relationship between I.Q. and the criteria was similar to that achieved by the D.T. scores and this was considered particularly important in the case of the Board Game (B.G.). In designing the latter, the writer attempted to incorporate demands on children's constructive, analytic and imaginative thinking ability, but had half expected conventional measures of intelligence to account for the differences in children's performance. The pilot work however indicated that this might not be the case and this was confirmed in the main experiment.

The results indicated that practical performance on such a task makes demands on a wide range of abilities, not confined to convergent thinking, and that divergent thinking abilities can make an important contribution. Forming a multiple predictor from I.Q. and two D.T. measures, Uses (fluency) and Circles (combined score), for example, the multiple correlation with boys' performance on the Board Game was 0.57, accounting for 32½% of the variance, 18% being contributed by the D.T. scores. It is unfortunate that having read the description and rationale of the B.G. in Chapter 5, a reader will not be in a position to try out on himself the principles involved. Studying children's, and even adults' approaches to the task, is in many ways more fascinating than their final solution, and although the method of approach was a contributory factor in the final performance category, and characteristic types of performance were described in Chapter 5, it is felt that the potential of the B.G. is not fully realised in this study.

Observing some children apparently siezing on properties of articles and functions of the material which were quite different from their habitual use, but appropriate to the problem in hand, and other children consistently failing to make appropriate associations, provided convincing examples of differences in children's thinking.

It is hoped to provide video-tape recordings of children's attempts at the B.G. for future teaching purposes, and the Board Game could also provide a basis for further research into children's thinking. Transcripts of the children's verbalisation of their thoughts and feelings at regular stages of their attempts, as utilized by Goor and Sommerfeld (1976), may provide useful insights into their creative behaviour in this respect.

For the purposes of this study however the Board Game demonstrated that D.T. abilities play a significant part in children's performance on such an exercise, and though the actual magnitude of the validity coefficients was not high, the writer maintains that they are of particular importance in illustrating the concurrent validity of divergent tests.

The results of the factorial investigation of construct validity gave further support to the existence of a cohesive dimension of divergent thinking abilities, which was relatively distinct from general convergent ability. It also demonstrated that divergent thinking could not be construed simply as a function of personality variables or attitudes. While a number of relationships were shown to exist between the latter and the tests of divergent thinking, convergent thinking, and the creative criteria, these were generally not very marked over the whole range of ability tested. In particular the absence of any relationship with extraversion enables Eysenck's (1967) suggestion, that divergent thinking is largely a matter of the latter, to be rejected.

With the evidence of the study as a whole supporting a construct of divergent thinking as a cognitive ability with reasonable reliability and stability over time, it is suggested that like other cognitive variables its application to other activities will, however, be mediated in numerous ways by variations in children's attitudes and personality. As indicated by Wallach and Kogan (1966) and Hudson (1966, 1968), particular subgroups chosen on the basis of contrasting or complementary abilities may reveal different relationships to personality and other variables, depending, for example, on whether high divergent thinking is accompanied by high or low intelligence, and vice versa. Both the concurrent and construct validity investigation indicated, however, that divergent thinking is not an entirely unidimensional activity, and more attention might be given to complementary rather than parallel tests of divergent thinking, verbal and figural abilities in particular showing some different associations with other variables. It is also possible that, having established a divergent thinking dimension, different types rather than levels of responses may, as Hudson (1968) and Hargreaves (1974, 1977) suggest, be worthy of study in their own right.

The amount of common variance extracted for the Uses Test and its comparable level of long-term stability with the other tests, suggested that its low test-retest reliability might have been an underestimate of its 'true' reliability. Its validity coefficients were generally higher than those for the Consequences Test, and with attention to motivation factors in any retest, it is suggested that the Uses Test and the Circles Test, representing verbal and figural aspects of divergent thinking respectively, would be a good complementary pair of tests in assessing divergent thinking abilities.

A number of other features have also been noted which might be confirmed with further investigation, and it may be possible to discover particular groupings of abilities and personality variables that may be particularly related to different types of learning and creative activity.

The final conclusion is that divergent thinking tests have emerged, after considerable scrutiny, as reasonably reliable measures of an intellectual ability which is relatively independent of intelligence, shows development and stability of relative ranking over time, and is positively related to creative behaviour.

It is consequently suggested that divergent thinking be recognised as one category in a broad conception of intellectual abilities such as proposed by Guilford (1956, 1967).

The identification of divergent thinking tests as 'creativity' tests is rejected, both theoretically and experimentally, but a limited amount of association between divergent thinking and creative behaviour has been demonstrated, and the evidence should provide some basic support for teachers who regard the fostering of divergent thinking abilities as a realistic and worthwhile aim.

REFERENCES

ABBOTT, S.D. (1972) The intermedia environment and its stimulation of the creative concomitant to reading success. Master's research paper. Old Dominican College.

American Educational Research Association (1955) *Technical Recommendations for Achievement Tests.* Washington D.C.: A.E.R.A.

American Psychological Association (1954) *Technical Recommendations for Psychological Tests and Diagnostic Techniques.* Washington: A.P.A.

American Psychological Association (1974) *Standards for Educational and Psychological Tests.* Washington: A.P.A.

ANASTASI, A. (1968) *Psychological Testing.* N.Y.: Macmillan.

ANASTASI, A. and SHAEFER, C.E. (1971) Note on the concepts of creativity and intelligence. *J.Creat.Behav.* 5, 2, 113-116.

ANDERSON, H.H. (1959) Creativity as personality development. Chapter 9 in Anderson H.H. (Ed.) *Creativity and its Cultivation.* N.Y.: Harper and Row.

ATKINSON, C.R.M. (1977) Personality and learning characteristics of intelligent 13-14 year old Millfield pupils in relation to their academic and extra-curricular achievement. Unpublished M.Ed. thesis. Univ. of Bath.

AUSUBEL, D.P. (1963) *The Psychology of Meaningful Verbal Learning.* N.Y.: Grune and Stratton.

AUSUBEL, D.P. (1968) *Educational Psychology: A Cognitive View.* N.Y.: Holt, Rinehart and Winston.

BAIRD, L.L. (1972) The Torrance Tests of Creative Thinking. Review in Buros, O.K. (Ed.) *7th Mental Measurements Yearbook.* N.J.: Gryphon Press.

BARKER LUNN, J.C. (1969) The development of scales to measure junior school children's attitudes. *Br.J.educ.Psychol.,* 39, 64-71.

BARKER LUNN, J.C. (1970) *Streaming in the Primary School.* Slough: N.F.E.R.

BARRON, F. (1953) Complexity-simplicity as a personality dimension. *J.abnorm.soc.Psychol.,* 48, 162-172.

BARTLETT, F.C. (1958) *Thinking: an Experimental and Social Study.* London: Allen and Unwin.

BEHRENS, R.R. (1974) The weave and warp of invention. *J.Creat.Behav.,* 8, 2, 81-92.

BENNETT, S.N. (1971) Review of *Creativity and Education* by Lytton, H. (1971) *Br.J.educ.Psychol.,* 41, 99.

BENNETT, S.N. and JORDAN, J. (1975) A typology of teaching styles in primary schools. *Br.J.educ.Psychol.,* 45, 20-28.

BERGER, R.M., GUILFORD, J.P., and CHRISTENSEN, P.R. (1957) A factor-analytic study of planning abilities. *Psychol. Monogr.,* 71, 6, No.435.

BERLYNE, D.E. (1965) *Structure and Direction in Thinking.* N.Y.: Wiley.

BERLYNE, D.E. (1972) Experimental aesthetics. In Dodewell, P.C. (Ed.) *New Horizons in Psychology 2*. Harmonsworth: Penguin, pp.107-125.

BIGGS, J.B., FITZGERALD, D., and ATKINSON, S.M. (1971) Convergent and divergent abilities in children and teachers' ratings of competence and certain classroom behaviours. *Br.J.educ.Psychol.*, 41, 277-286.

BINET, A. and HENRI, V. (1896) La psychologie individuelle. *Annêe Psychol.*, 2, 411-465.

BISH, G.C. (1964) A study of the relationships of intelligence, achievement, creativity, anxiety and confidence among pupils in a suburban elementary school. Doctoral dissertation. Washington. George Washington Univ.

BRODBECK, M. (1963) Logic and scientific method in research on teaching. In Gage, N.L. (Ed.) *Handbook of Research in Teaching*. Chicago: Rand McNally, pp.44-93.

BROWN, F.G. (1976) *Principles of Educational and Psychological Testing* (2nd Edn.). N.Y.: Holt, Rinehart and Winston.

BRUNER, J.S. (1960) *The Process of Education*. Cambridge, Mass: Harvard Univ. Press.

BRUNER, J.S. (1961) The act of discovery. *Harvard Educ.Review*, 31, 21-32.

BRUNER, J.S., GOODNOW, J.J., and AUSTIN, G.A. (1956) *A Study of Thinking*, N.Y.: Wiley.

BUROS, O.K. (1972) (Ed.) *7th Mental Measurements Yearbook*. N.Y.: Gryphon.

BURT, C. (1940) *The Factors of the Mind*. London: ULP.

BURT, C. (1949) The structure of the mind: A review of the results of factor analysis. *Br.J.educ.Psychol.*, pp.19, 100-111, 176-199.

BURT, C. (1952) Tests of significance in factor analysis. *Br.Journal Psych.* (Statistical Section) pp.5, 109-133.

BURT, C. (1962) The psychology of creative ability: Critical notice of *Creativity and Intelligence* by Getzels and Jackson. *Br.J.educ.Psychol.*, 32 pp.292-298.

BUTCHER, H.J. (1968) *Human Intelligence*. London: Methuen.

BUTCHER, H.J. (1969) The structure of abilities, interests and personality in 1,000 Scottish school children. *Br.J.educ. Psychol.*, Vol.39, 154-165.

BUTCHER, H.J. (1972) Divergent thinking and creativity. In Wall, W.D. and Varma, V.P. (Eds.) *Advances in Educational Psychology*, Vol.1 pp.85-95. London: ULP.

CAMPBELL, D.T., (1960) Recommendations for APA test standards regarding, construct trait, or discriminant validity. *Amer.Psychol.*, 15, 546-553.

CATTELL, J.McK. (1890) Mental tests and measurements. *Mind*, 15, 373-380.

CATTELL, R.B. (1952) *Factor Analysis*. N.Y.: Harper.

CATTELL, R.B. (1964) Validity and reliability: a proposed more basic set of concepts. *J.Educ.Psychol.*, 55, 1, 1-22.

CATTELL, R.B. (1971) *Abilities: Their Structure, Growth, and Action.* Boston: Houghton Mifflin.

CATTELL, R.B., and BUTCHER, H.J. (1968) *The Prediction of Achievement and Creativity.* N.Y.: Bobbs-Merrill.

CATTELL, R.B., and DREVDAHL, J.E. (1955) A comparison of the personality profile of eminent researchers with that of eminent teachers and administrators and the general population. *Br.J.Psychol.*, 46, pp.248-261.

CHANNON, C.E. (1974) The effects of regime on divergent thinking scores. *Br.J.educ.Psychol.*, 44, 89-91.

CHILD, D. (1968) Convergent and divergent thinking and arts/science. *Abst.Br.Psychol.Soc.*, Education Section Annual Conference, 1968.

CHILD, D. (1970) *Essentials of Factor Analysis.* London: Holt Rinehart.

CHILD, D., and SMITHERS, A.G. (1971) Some cognitive and affective factors in subject choice. *Res.in.Educ.*, 5, 1-9.

CHILD, D., and SMITHERS, A.G. (1973) An attempted validation of the Joyce-Hudson scale of convergence and diverence. *Br.J.educ.Psychol.*, 43, 57-62.

CROCKENBERG, S.B. (1972) Creativity tests: A boon or boondoggle for education. *Rev.Ed.Res.*, 42, 27-46.

CRONBACH, L.J. (1968) Intelligence? Creativity? A parsimonious reinterpretation of the Wallach-Kogan data. *Amer.Educ. Res.J.*, 5, 4, 491-511.

CRONBACH, L.J. (1970) *Essentials of Psychological Testing* (3rd. edn.). N.Y.: Harper and Row.

CRONBACH, L.J. (1971) Test validation. In Thorndike, R.L., (Ed.) *Educational Measurement* (2nd edn.). Washington: American Council on Education, pp.443-507.

CRONBACH, L.J., and MEEHL, P.E. (1955) Construct validity in psychological tests. *Psychol.Bull.*, 52, 281-302. Reprinted in Noll, V.H. *et al* (Eds.)(1972) *Introductory Readings in Educational Measurement.* Boston: Houghton Mifflin, pp.90-121.

CROPLEY, A.J. (1964) Differentiation of abilities, socioeconomic status and the WISC. *J.consult.Psychol.*, 28, 512-517.

CROPLEY, A.J. (1967) *Creativity.* London: Longmans.

CROPLEY, A.J. (1968) A note on the Wallach-Kogan tests of creativity. *Br.J.educ.Psychol.*, Vol.38, 197-201.

CROPLEY, A.J. (1972) A five-year longitudinal study of the validity of creativity tests. *Dev.Psychol.*, 6, 119-124.

CROPLEY, A.J., and CLAPSON, L. (1971) Long term test-retest reliability of creativity tests. *Br.J.educ.Psychol.*, 41, 206-208.

CROPLEY, A.J., and MASLANY, G.W. (1969) Reliability and factorial validity of the Wallach-Kogan creativity tests. *Br.J.Psychol.*, 60, 3, 395-398.

CROSS, P.C., CATTELL, R.B., and BUTCHER, H.J. (1967) The personality pattern of creative artists. *Br.J.educ.Psychol.*, 37, 3, 292-299.

CHRISTIE, T. (1970) Cognitive bias, university faculty and degree of success. *Abst.Br.Psychol.Soc.* Education Section, Annual Conference, 1970.

CHRUTCHFIELD, R.S. (1964) Conformity and creative thinking. In Gruber, Terrell and Wertheimer (Eds.) *Contemporary Approaches to Creative Thinking.* N.Y.: Atherton Press, pp.120-140.

DACEY, J., and MADAUS, G.F. (1969) Creativity: definitions, explanations and facilitation. *Irish J.of Educ.,* iii, 1, 55-69.

DAUW, D.C. (1965) Life experiences, vocational needs and choices of original thinkers and good elaborators. Doctoral dissertation. Minneapolis: Univ. of Minnesota.

DEARDEN, R.F. (1968) *The Philosophy of Primary Education: an Introduction.* London: Routledge and Kegan Paul.

DE BONO, E. (1967) *The Use of Lateral Thinking.* London: Jonathan Cape.

DE BONO, E. (1969) Information processing and new ideas - lateral and vertical thinking. *J.Creat.Behav.,* 3, 3, 159-171.

DE BONO, E. (1971) *Practical Thinking.* London: Jonathan Cape.

DE BONO, E. (1973) *CORT Thinking: Teachers' Notes.* Blandford, Dorset: Direct Educational Services Ltd.

DALBEC, E. (1966) Creative development over a three-year period in a Catholic Liberal Arts College. Master's research report. Minneapolis: Univ. of Minnesota.

DAVIDSON, H.H. and GREENBERG, J.W. (1967) *School Achievers from a Deprived Background.* Project No.2805, No.OE-5-10-132. City College, City Univ. of N.Y., May 1967.

DEBNEY, B. (1969) Testing creativity. *Ed.Review,* 22, 196-208.

DENTLER, R.A., and MACKLER, B. (1964) Originality: Some social and personal determinants. *Behav.Sci.,* 9, 1-7.

DERNER, G.F., ABORN, M., and CANTER, A.H. (1950) The reliability of the Wechsler-Bellevue subtests and scales. *J.consult. Psychol.,* 14, 172-179.

DEWEY, J. (1933) *How we Think.* Boston: Heath.

DEWING, K. (1970) The reliability and validity of selected tests of creative thinking in a sample of seventh grade West Australian children. *Br.J.educ.Psychol.,* 40, 35-42.

DREVDAHL, J.E. (1956) Factors of importance for creativity. *J.Clin.Psychol.,* 12, 21-26.

DREVER, J. (1952) *Dictionary of Psychology.* Harmondsworth: Penguin.

DUDERK, S.Z. (1974) Creativity in young children- attitude or ability? *J.Creat.Behav.,* 8, 4, 282-292.

DUNKER, K. (1945) On problem-solving. *Psychol.Monogr.* No.270.

Educational Testing Service (1961) *Selecting an Achievement Test: Principles and Procedures.* Princeton, N.J.: ETS.

ENTWISTLE, N.J. (1968) Academic motivation and school attainment. *Br.J.educ.Psychol.,* 38, 181-8.

ENTWISTLE, N.J. (1972) Personality and academic attainment. In Butcher, H.J., and Pont, H.B., *Educational Research in Britain*, Vol.3. London: ULP, pp.89-111.

ENTWISTLE, N.J., and CUNNINGHAM, S. (1968) Neuroticism and school attainment - a linear relationship? *Br.J.educ. Psychol.*, 38, 123-32.

EVANS, K.M. (1965) *Attitudes and Interests in Education.* London: RKP.

EYSENCK, H.J. (1953) The logical basis of factor analysis. *Amer.Psychol.*, 8, 105-114.

EYSENCK, H.J. (1967) Intelligence assessment: A theoretical and experimental approach. *Br.J.educ.Psychol.*, Vol.37, 81-98.

EYSENCK, H.J. (1972) Personality and learning. In Wall, W.D., and Varma, V.P. (Eds.) *Advances in Educational Psychology* Vol.1. London: ULP, pp.96-112.

EYSENCK, H.J. and COOKSON, D. (1969) Personality in Primary school children. I Ability and achievement, II Teachers' ratings. *Br.J.educ.Psychol.*, Vol.39, 109-130.

EYSENCK, S.B.G. (1973) *Manual of the Junior Eysenck Personality Inventory.* London: ULP.

EYSENCK, S.B.G., NIAS, D.K.B., and EYSENCK, H.J. (1971) The interpretation of children's lie scale scores. *Br.J.educ.Psychol.*, 41, 23-31.

EYSENCK, S.B.G., and EYSENCK, H.J. (1973) Test-retest reliabilities of a new personality questionnaire for children. *Br.J.educ.Psychol.*, 43, 126-130.

FEE, F. (1968) An alternative to Ward's factor analysis of Wallach and Kogan's 'Creativity' correlations. *Br.J.educ. Psychol.*, Vol.38, 319-321.

FESTINGER, L. (1957) *A Theory of Cognitive Dissonance.* Stanford U.P.

FISHER, S. (1950) Patterns of personality rigidity and some of their determinants. *Psychol. Monogr.*, 64, 1-48.

FLEMING, E.S. and WEINTRAUB, S. (1962) Attitudinal rigidity as a measure of creativity in gifted children. *J.Educ. Psychol.*, 53, 81-85.

FRANKLIN, B.D., and RICHARDS, P.N. (1977) Effects on children's divergent thinking abilities of a period of direct teaching for divergent production. *Br.J.educ.Psychol.*, 47, 66-70.

FREEMAN, J., BUTCHER, H.J., and CHRISTIE, T. (1971) *Creativity* (2nd edn.). London: Soc.Res.Higher Educ.

FRENKEL-BRUNSWIK, E., (1950) In Adorno, T.W., Frenkel-Brunswick, E., *et al. The Authoritarian Personality.* N.Y.: Harper.

FREUD, Sigmund (1949) *The Unconscious.* Collected Papers IV. London: Hogarth.

FRUCHTER, B. (1954) *Introduction to Factor Analysis.* N.Y.: Van Nostrand.

GAGNÉ, R.M. (1964) Problem Solving. In Melton, A.W. (Ed.) *Categories of Human Learning.* London: Academic Press.

GAGNÉ, R.M. (1970) *The Conditions of Learning.* London: Holt, Rinehart and Winston.

GALLAGHER, J.J. (1964) Productive thinking. In Hoffman, M.L., and Hoffman, L.W. (Eds.) *Child Devel.Research.* Russell Sage Foundation.

GALTON, Francis (1869) Hereditary genius: an inquiry into its laws and consequences. N.Y.: Macmillan. Chapter 3 is reprinted in Jenkins, J.J. and Patterson, D.G. (1961) *Studies in Individual Differences.* N.Y.: Appleton.

GARDNER, D.E.M. (1950) *Long Term Results of Infant School Methods.* London: Methuen.

GARDNER, D.E.M. (1966) *Experiment and Tradition in Primary Schools.* London: Methuen.

GARRETT, H.E. (1966) *Statistics in Psychology and Education* (6th edn.) London: Longmans.

GARWOOD, D.S. (1964) Personality factors related to creativity in young scientists. *J.Abnorm.Soc.Psychol.,* 68, 413-419.

GETZELS, J.W. and JACKSON, P.W. (1962) *Creativity and Intelligence.* N.Y.: Wiley.

GHISELIN, B. (Ed.)(1955) *The Creative Process.* N.Y.: Mentor.

GHISELIN, B. (1957) Ultimate criteria for two levels of creativity. In Taylor, C.W. (1958) *The Second (1957) Univ. of Utah Research Conference,* pp.141-155.

GOLANN, S.E. (1961) The creativity motive. Unpublished Ph.D. thesis. Univ. of North Carolina.

GOLANN, S.E. (1962) The creativity motive. *J.Person.,* 30(4), 588-600.

GOLANN, S.E. (1963) Psychological study of creativity. *Psychol.Bull.,* 60, 6, 548-565.

GOLDBERG, M.L. (1965) *Research on the Talented.* N.Y.: Bureau of Publications, Teachers College, Columbia Univ.

GOLDMAN, R.J. (1967) The Minnesota tests of creative thinking. In Mooney, R.L. and Razik, T.A. (Eds.) *Explorations in Creativity,* N.Y.: Harper and Row, pp.267-280.

GOLDMAN, R.J., and CLARKE, D.F. (1967) The Minnesota tests of creative thinking - a note on scorer reliability in follow-up studies with English primary school children. *Br.J.educ.Psychol.,* 37, 115-117.

GOOR, A., and SOMMERFELD, R.E. (1975) A comparison of problem-solving processes of creative students and noncreative students. *J.Educ.Psychol.,* 4, 495-505.

GILCHRIST, M., *The Psychology of Creativity.* Carlton, Australia: Melbourne Univ.Press.

GUILFORD, J.P. (1950) Creativity. *Amer.Psychol.,* 5, 444-454.

GUILFORD, J.P. (1951) with WILSON, R.C., CHRISTENSEN, P.R., and LEWIS, D.J. A factor-analytic study of creative thinking. I. Hypotheses and description of tests. *Rep.Psychol.Lab.No.4,* Univ.S.Calif.

GUILFORD, J.P. (1952) Some recent findings on thinking abilities and their implications. *Inform.Bull.*, 3, 48-61.

GUILFORD, J.P. (1954)(*et al*) *Report Psychol.Lab.No.14:*Univ. of S.Calif.

GUILFORD, J.P. (1956) The structure of intellect. *Psychol. Bull.*, Vol.53, 4, 267-293.

GUILFORD, J.P. (1958) New frontiers of testing in the discovery and development of human talent. In Guilford, J.P. (1968) *Intelligence, Creativity and their Educational Implications.* San Diego: Knapp, pp.3-13.

GUILFORD, J.P. (1959, 1968) Three faces of Intellect. *Amer. Psychol.*, 14, 469-479. Reprinted (1968) in Guilford, J.P., *Intelligence, Creativity and their Educational Implications.* San Diego: Knapp.

GUILFORD, J.P. (1959, 1973) Traits of creativity. In Anderson, H.H. (Ed.) *Creativity and its Cultivation*, Harper, pp.142-61. Reprinted (1973) in Vernon P.E. (Ed.) *Creativity.* Harmondsworth: Penguin, pp.167-188.

GUILFORD, J.P. (1962) Factors that aid and hinder creativity. *Teach.Coll.Rec.*, 63, 380-392.

GUILFORD, J.P. (1965) Intelligence: 1965 model. *Amer.Psychol.*, 1966, 21, 20-26. Reprinted in Guilford, J.P. (1968) *Intelligence, Creativity and their Educational Implications.* San Diego: Knapp.

GUILFORD, J.P. (1967) *The Nature of Human Intelligence.* New York: McGraw Hill.

GUILFORD, J.P. (1967b) Measurement of creativity. In Mooney, R.L. and Razik, T. (Eds.) *Explorations in Creativity*, N.Y.: Harper and Row, pp.281-287.

GUILFORD, J.P. (1970) Creativity: retrospect and prospect. *J.Creat.Behav.*, 4, 149-168.

GUILFORD, J.P. (1975) Relation of divergent-production abilities to verbal and non-verbal I.Q.'s. *J.Mult.Expt.Pers. Clin.Psychol.*, 1(4), 278-287.

GUILFORD, J.P. (1976) Aptitude for creative thinking: one or many? *J.Creat.Behav.*, 10, 165-169.

GUILFORD, J.P. (with FRUCHTER, B.)(1973) *Fundamental Statistics in Psychology and Education.* Tokyo: McGraw-Hill Kogakusha.

GUILFORD, J.P. and HOEPFNER, R. (1966) Structure-of-intellect factors and their tests 1966. *Rep.Psychol.Lab.No.36*, Univ.S.Calif.

GUILFORD, J.P., WILSON, R.C., and CHRISTENSEN, P.R. (1952) A factor-analytic study of creative thinking. II. Administration of tests and analysis of results. *Rep.Psychol. Lab.No.8*, Univ.S.Calif.

GUTMAN, H. (1967) The biological tests of creativity. In Mooney, R.L. and Razik, T. (Eds.) *Explorations in Creativity.* N.Y.: Harper and Row, pp.3-32.

HADAMARD, Jacques (1949) *The Psychology of Invention in the Mathematical Field.* Princeton: Princeton Univ.Press.

HADDON, F.A., and LYTTON, H. (1968) Teaching approach and the development of divergent thinking abilities in primary schools. *Br.J.educ.Psychol.*, 38, 171-180.

HADDON, F.A., and LYTTON, H. (1971) Primary education and divergent thinking abilities - four years on. *Br.J.educ. Psychol.*, 41, 136-147.

HAGENDER, H. (1967) Influences of creative writing experiences on general creative development. Master's research paper. Minneapolis: Univ.of Minnesota.

HALPIN, G., and HALPIN, G. (1973) The effect of motivation on creative thinking abilities. *J.Creat.Behav.*, 7, 1, 51-53.

HALPIN, G., HALPIN, G., and TORRANCE, E.P. (1974) Relationships between creative thinking abilities and a measure of the creative personality. *Educ.Psychol.Meast.*, 34, 74-82.

HAMMAKER, M.K., SHAFTO, M., and TRABASSO, T., (1975) Judging creativity: a method for assessing how and by what criteria it is done. *J.Educ.Psychol.*, 67, 478-483.

HARGREAVES, D.J. (1974) Situational influences on divergent thinking. *Br.J.educ.Psychol.*, 44, 84-88.

HARGREAVES, D.J. (1974) Psychological testing: current perspectives and future developments. *Ed.Rev.*, 27, 1, 26-33.

HARGREAVES, D.J. (1977) Sex roles in divergent thinking. *Br.J. educ.Psychol.*, 47, 25-32.

HARGREAVES, D.J., and BOLTON, N. (1972) Selecting creativity tests for use in research. *Br.J.Psychol.*, 63, 3, 451-462.

HARGREAVES, H.L. (1927) The 'faculty' of imagination. *Br.J. Psychol.Monogr.*, Supplement No.10.

HARMAN, H.H. (1967) *Modern Factor Analysis*. Chicago: Univ.of Chicago Press.

HASAN, P., and BUTCHER, H.J. (1966) Creativity and intelligence. A partial replication with Scottish children of Getzels and Jackson's study. *Br.J.of Psychol.*, Vol.57, 129-135.

HEIM, A.W. (1968) *Manual. AH5 Group Test of High-Grade Intelligence*. Windsor, Berks: N.F.E.R.

HEIM, A.W. (1970) *Intelligence and Personality*. Harmondsworth, Middlesex: Penguin.

HOLLAND, J.L. (1972) Torrance tests of creative thinking. Review in Buros, O.K. *7th Mental Measurements Yearbook*. N.Y.: Gryphon Press, pp.840-841.

HOLT, P.N. (1969) A study of children's responses to open-ended problems. Unpublished Ph.D. thesis. Univ.of Southampton.

HUDSON, L. (1966) *Contrary Imaginations*. London: Methuen.

HUDSON, L. (1968) *Frames of Mind*. London: Methuen.

HUDSON, L. (1972) *The Cult of the Fact*. London: Jonathan Cape.

HUMPHREY, G. (1951) *Thinking*. London: Methuen.

HUMPHREYS, L.G. (1962) The organisation of human abilities. *Amer.Psychol.*, 17, 475-483.

KAMIN, L.T. (1976) Heredity, Intelligence, Politics and Psychology I and II. In Block, N.J., and Dwarkin, G. (Eds.) *The I.Q. Controversy.* London: Onartel Bks., pp.242-264 and 374-382.

KAGAN, J. (1966) Developmental studies in reflection and analysis. In Kidd, A.H., and Rivoire, J.E. (Eds.) *Perceptual Development in Children.* N.Y.: Intern.Universities Press. Reprinted in Cashdan, A., and Whitehead, J. (Eds.) *Personality Growth and Learning.* Open University Source Book. London: Longmans, pp.36-56.

KAISER, H.F. (1958) The varimax criterion for analytic rotation in factor analysis. *Psychomotrika.,* 23, 187-200.

KARVÉ, B.D. (1929) An experimental investigation of 'fluency' of imagination and association in school children. Unpublished M.Ed. Thesis, Univ.of Leeds.

KELLMER PRINGLE, M.L., and McKENZIE, J.R. Teaching method and rigidity in problem solving. *Br.J.educ.Psychol.,* 35, 50-59.

KERR, J.F. (1966) Science teaching and social change. *School Sci.Rev.,* 47, 301-309.

KLEIN, BARR and WOLITZKY (1967) Creativity. *Annual Review of Psychol.,* pp.536-539.

KNELLER, G.F. (1966) *The Art and Science of Creativity.* N.Y.: Holt, Rinehart and Winston.

KOESTLER, A. (1964) *The Act of Creation.* London: Hutchinson.

KOGAN, N., and PANKOVE, E. (1974) Long-term predictive validity of divergent thinking tests: some negative evidence. *J.Educ.Psychol.,* 66, 6, 802-810.

KRETSCHMER, E. (1931) *The Psychology of Men of Genius.* London: Kegan Paul.

KUBIE, L.S. (1965) Blocks to creativity. *Int.Sc.Technol.,* pp.67-78.

KUDER, G.F., and RICHARDSON, M.W. The theory of the estimation of test reliability. *Psychometrika,* 2, 151-160.

LANG, R.J., and RYBA, K.A. (1976) The identification of some creative thinking parameters common to the artistic and musical personality. *Br.J.educ.Psychol.,* 46, 267-279.

LEACH, P.J. (1964) Teaching tolerance. *Int.Rev.Educ.,* 10, 190-204.

LEACH, P.J. (1967) A critical study of the literature concerning rigidity. *Br.J.Soc.Clin.Psychol.,* 6, 11-22. Reprinted in Warr, P.D. (Ed.) *Though and Personality.* Harmondsworth: Penguin.

LEWIS, D.G. (1967) *Statistical Methods in Education.* London: ULP.

LEWIS, D.G. (1974) *Assessment in Education.* London: ULP.

LIEBERMANN, J.N. (1965) Playfulness and divergent thinking: an investigation of their relationship at the Kindergarten level. *J.Genet.Psychol.,* 107, 219-224.

LONG, B.H. *et al* (1965) *Self-social Correlates of Originality in Children.* Report. Newark: Univ.of Delaware. Mimeographed.

LONG, B.H., and HENDERSON, E.H. (1964) *Opinion Formation and Creativity in Elementary Children.* Report. Newark: Univ.of Delaware. Mimeographed.

LOVELL, K., and SHIELDS, J.B. (1968) Some aspects of a study of the gifted child. *Br.J.educ.Psychol.,* Vol.37, 201-208.

LUCAS, A.M. (1971) Creativity, discovery and inquiry in science education. *Aust.J.Educ.,* 15, 2, 185-196.

LUCHINS, A.S. (1942) Mechanisation in problem solving: the effect of Rinstellung. *Psychol.Monogr.,* No.248.

LUCHINS, A.S. (1951) On an approach to social perception. *J.Personal.,* 19, 64-84.

LYNN, R., and GORDON, I.E. (1961) The relation of neuroticism and extraversion to intelligence and educational attainment. *Br.J.educ.Psychol.,* 31, 194-203.

LYTTON, H., and COTTON, A.C. (1969) Divergent thinking abilities in secondary schools. *Br.J.educ.Psychol.,* 39, 188-90.

LYTTON, H. (1971) *Creativity and Education.* London: R.K.P.

MACKAY, C.K., and CAMERON, M.B. (1968) Cognitive bias in Scottish first-year science and arts undergraduates. *Br.J.educ.Psychol.,* 38, 315-18.

MACKLER, B., and SHONTZ, F.C. (1965) Creativity: theoretical and methodological considerations. *Psychol.Rec.,* 15, 217-238.

MACKINNEN, D.N. (1962) The nature and nurture of creative talent. *Amer.Psychologist.,* 17, 484-495.

MCKELLAR, P. (1957) *Imagination and Thinking.* London: Cohen and West.

MCNEMAR, Q. (1951) The factors in factoring behaviour. *Psychometrika.,* 16, 352-359.

MCNEMAR, Q. (1964) Lost: our intelligence? Why? *Amer. Psychologist.,* 19, 871-882.

MAIER, N.R.F. (1930) Reasoning in humans: I. On direction. *J.comp.Psychol.,* 10, 115-143.

MAIER, N.R.F. (1931) Reasoning in humans: II. The solution of a problem and its appearance in consciousness. *J.comp. Psychol.,* 12, 181-94. Extract in Watson, P.E., and Johnson-Laird, P.N. (1968) *Thinking and Reasoning.* Harmondsworth: Penguin.

MAIER, N.R.F. (1945) Reasoning in humans: IV. The mechanisms of equivalent stimuli and reasoning. *J.exper.Psychol.,* 35, 349-360.

MARSH, R.W. (1964) A statistical re-analysis of Getzels and Jackson's data. *Br.J.educ.Psychol.,* 34, 91-93.

MATTHEWS, G. (1969) Mathematics in the middle years. In, Schools Council, *The Middle Years of Schooling from 8 to 13.* Working Paper 22. London:H.M.S.O.

MEDNICK, S.A. (1962) The associative basis of the creative process. *Psychol.Review,* 69, 220-232.

MICHAELIS, W., and EYSENCK, H.J. (1971) The determination of personality inventory factor patterns and intercorrelations by changes in real life motivation. *J.Gen.Psychol.*, 118, 223-234.

MILGRAM, R.M., and MILGRAM, N.A. (1976) Creative thinking and creative performance in Israeli students. *J.Educ.Psychol.*, 68, 3, 255-259.

MOONEY, R.L. (1967) Creation in the classroom setting. In Mooney, R.L. and Razik, T. *Explorations in Creativity*. N.Y.: Harper and Row, pp.206-215.

MORSE, W.C., and WINGO, G.M. (1962) *Psychology and Teaching*. Chicago: Scott Foresman.

MOUSTAKAS, C. (1966) *The Authentic Teacher*. Cambridge, Mass.: Howard Doyle.

NELSON, J.F. (1963) The construction of a scale of teacher judgement of pupil creativity. Master's research paper. Duluth: Univ.of Minnesota.

NIE, N.H. *et al* (1975) *SPSS - Statistical Package for the Social Sciences*. N.Y.: McGraw-Hill.

Nuffield Foundation (1964a) *Bulletin No.1*. London: Nuffield Foundation Mathematics Teaching Project.

Nuffield Foundation (1964b) *Progress Report*. London: Nuffield Foundation Science Teaching Project.

NUTTALL, D.L. (1971) Modes of thinking and their measurement. Unpublished Ph.D. thesis. University of Cambridge.

NUTTALL, D.L. (1972) Convergent and divergent thinking. In Butcher, H.J., and Pont, H.B. (Eds.) *Educational Research in Britain*, Vol.3, pp.112-129. London: ULP.

OGILVIE, E. (1974) Creativity and curriculum structure. *Ed.Res.*, 16, 2, 126-132.

OLKIN, S.P. (1967) The relationship of creative problem-solving to creative thinking in junior high school students. Master's research report. Queens College (N.Y.).

OSBORN, A.F. (1957) *Applied Imagination*. N.Y.: Scribners.

PASK, G. (1969) Strategy, competence and conversation as determinants of learning. *Prog.Learn.Educ.Technol.*, 6, 250-67.

PATRICK, Catherine (1938) Scientific thought. *J.Psychol.*, 5, 55-83.

PETERS, R.S. (1969)(Ed.) *Perspectives on Plowden*. London: RKP.

PILLINER, A.E.G. (1968) Examinations. In Butcher, H.J. (Ed.) *Educational Research in Britain*. London: ULP.

PLASS, H., MICHAEL, J.J., and MICHAEL, W.B. (1974) The factorial validity of the 'Torrance Tests of Creative Thinking' for a sample of 111 sixth-grade children. *Educ.Psychol. Meast.*, 34, 413-414.

Plowden Report (1967) *Children and their Primary Schools*. D.E.S. London: H.M.S.O.

POINCARÉ, Henri (1906,1968) Mathematical creation. Lecture to the Psychological Society. Paris. Reprinted (1968) in *Mathematics in the Modern World*. Scientific American.

POLE, K.E. (1969) A study of techniques for measuring creative ability and the development and validation of creativity tests for highly intelligent subjects. Unpublished B.Litt. thesis. University of Oxford.

RAINA, M.K. (1971) Research developments in creativity in India. *J.Res.Devel.Educ.*, 4 (3), 118-128.

REGAN, C.M. (1976) Social class and creativity. Unpublished M.Ed. dissertation. Univ.of Bath.

RICHARDS, P.N. (1970) A study of some effects on children's creative thinking of the discovery approach to mathematics in the primary school. Unpublished M.Ed. thesis. Univ.of Durham.

RICHARDS, P.N., and BOLTON, N. (1971) Type of mathematics teaching, mathematical ability and divergent thinking in junior school children. *Br.J.educ.Psychol.*, 41, 32-37.

RICHARDS, R.L. (1976) A comparison of selected Guilford and Wallach-Kogan tests in conjunction with measures of intelligence. *J.Creat.Behav.*, 10, 151-164.

RIPPE, D.D. (1953) Application of a large sampling criterion to some sampling problems in factor analysis. *Psychometrika.*, 18, 191-205.

ROE, A. (1953) *Making a Scientist.* N.Y.: Dodd, Mead & Co.

ROGERS, C.R. (1954) Toward a theory of creativity. *ETC: A Review of General Semantics*, 11, pp.249-60. Reprinted in Verna, P.E. (1970) *Creativity.* Harmondsworth: Penguin, pp.137-151.

ROSSMAN, J. (1971) *The Psychology of the Inventor.* Washington D.C.: Inventors Publishing Co.

RUGG, H. (1963) *Imagination.* Harper and Row.

SATTERLY, D.J., and BRIMER, M.A. (1971) Cognitive styles and school learning. *Br.J.educ.Psychol.*, 41, 294-303.

Schools Council (1966) Multiple Marking of English Compositions. Exam. Bulletin No.12. London: H.M.S.O.

SHAPIRO, R.J. (1968) Creative Research Scientists. *Psychol. Afric.Monogr.*, 4, 37-45. Reprinted in Vernon, P.E. (1973) *Creativity.* Harmondsworth: Penguin, pp.257-269.

SHAEFER, C.E. (1970) Evaluation of a program for developing creative thinking in teachers and children at the 4th and 5th grade levels. Bronx N.Y.: Creativity Centre, Fordham Univ.

SECHREST, L. (1963) Incremental validity: A recommendation. *Educ.Psychol.Meast.*, 23, 153-158.

SHARPLES, D. (1969) Children's attitudes towards Junior School activities. *Br.J.educ.Psychol.*, 39, 72-77.

SHOUKSMITH, G. (1970) *Intelligence, Creativity and Cognitive Style.* London: Batsford.

SHULMAN, L. (1966) Review of Wallach and Kogan (1966), Modes of thinking in young children. *Am.Ed.Res.J.*, 3: 305-309.

SMITHERS, A.G., and CHILD, D. (1974) Convergers and divergers: different forms of neuroticism? *Br.J.educ.Psychol.*, 44, 304-306.

SPEARMAN, C. (1930) *The Creative Mind.* London: Nisbet & Co.

SPIELMAN and GAW (1926) Special abilities. In *A Study in Vocational Guidance.* Nat.Inst.of Indus.Psychol., Ch.5.

STANLEY, J.C. (1971) Reliability. In Thorndike, R.L. (Ed.) Educational Measurement. (2nd edn.). Washington: American Council on Education, pp.356-442.

STARR, J.W., and NICHOLL, C. (1975) Creativity and Achievement in Nuffield Physics. *Br.J.educ.Psychol.*, 45, 322-326.

STEIN, M.I. (1967) Creativity and culture. In Mooney, R.L., and Razik, T. (Eds.) *Explorations in Creativity.* N.Y.: Harper and Row, pp.109-119.

STORR, A. (1972) *The Dynamics of Creation.* London: Secker and Warburg. (Also in Penguin Books, 1976).

STENHOUSE, D. (1973) *The Evolution of Intelligence.* London: Allen and Unwin.

TAYLOR, C.W. (1958) The second (1957) University of Utah research conference on the identification of creative scientific talent. Salt Lake City: Univ.of Utah Press.

TAYLOR, C.W. (1964) Introduction. In Taylor, C.W. (Ed.) *Creativity: Progress and Potential.* N.Y.: McGraw Hill, pp.1-14.

TAYLOR, C.W., SMITH, W.R., and GHISELIN, B. (1963) The creative and other contributions of one sample of research scientists. In Taylor, C.W. and Barron, F. (Eds.) *Scientific Creativity: Its Recognition and Development.* N.Y.: Wiley, pp.53-76.

TAYLOR, D.W. (1963) Variables related to creativity and productivity among men in two research laboratories. In Taylor, C.W., and Barron, F. (Eds.) *Scientific Creativity: Its Recognition and Development.* N.Y.: Wiley.

TAYLOR, I.A. (1959) The nature of the creative process. In Smith, P. (Ed.) *Creativity.* N.Y.: Hastings Rouse, pp.51-82.

TERMAN, L.M. (1906) Genius and stupidity - a study of some of the intellectual processes of seven 'brighter' and seven 'stupid' boys. *Pedag.Sem.*, 13, 307-373.

TERMAN, L.M. (1925) *Mental and Physical Traits of a Thousand Gifted Children. Genetic Studies of Genius I.* Stanford: Stan.U.P.

TERMAN, L.M., and ODEN, M.H. (1947) *The Gifted Child Grows Up. Genetic Studies of Genius IV.* Stanford: Stan.U.P.

TERMAN, L.M., and ODEN, M.H. (1959) *The Gifted Group at Mid-Life. Genetic Studies of Genius V.* Stanford: Stan.U.P.

THOMSON, G.H. (1950) *The Factorial Analysis of Human Ability.* London: L.U.P.

THOMSON, R. (1959) *The Psychology of Thinking.* Harmondsworth: Penguin.

THORNDIKE, R.L. (1951) Reliability. In Lindquist, E.F. (Ed.) *Educational Measurement.* Washington: American Council on Education, pp.560-620.

THORNDIKE, R.L. (1966) Some methodological issues in the study of creativity. Proc. of 1962 Invitation Conference on testing problems. Princeton: ETS.

THORNDIKE, R.L., and HAGAN, E.P. (1969) *Measurement and Evaluation in Psychology and Education* (3rd edn.). N.Y.: Wiley.

THURSTONE, L.L. (1952) Creative talent. In Thurstone, L.L. (Ed.) *Applications of Psychology.* N.Y.: Harper and Row, pp.18-37.

TORRANCE, E.P. (1959) Explorations in creative thinking in the early school years. Res.Monogr., 59/12. Univ.of Minnesota. Reprinted in Taylor, C.W. (1959) *The Third (1959) University of Utah Research Conference on the Identification of Creative Talent.* Salt Lake City: Univ.of Utah Press.

TORRANCE, E.P. (1962) *Guiding Creative Talent.* Englewood Cliffe, N.J.: Prentice-Hall.

TORRANCE, E.P. (1962b) Who is the underachiever? *NEA Journal*, 51, 15-17.

TORRANCE, E.P. (1963) *Education and the Creative Potential.* Minneapolis: Univ.of Minnesota Press.

TORRANCE, E.P. (1964) Education and creativity: In Taylor, C.W. (Ed.) *Creativity: Progress and Potential.* N.Y.: McGraw-Hill, pp.49-128.

TORRANCE, E.P. (1965) *Rewarding Creative Behaviour.* Englewood Cliffe, N.J.: Prentice-Hall.

TORRANCE, E.P. (1966) *Torrance Tests of Creative Thinking: Norma-Technical Manual.* Research Edition. Princeton, N.J.: Personnel Press.

TORRANCE, E.P. (1967) Non-test ways of identifying the creatively gifted. In Gowan, C.J. *et al* (Eds.) *Creativity: Its Educational Implications.* N.Y.: Wiley.

TORRANCE, E.P. (1969) Prediction of adult creative achievement among High School seniors. *Gifted Child Quarterly*, 13, 871-81.

TORRANCE, E.P. (1971) Long-range prediction studies and international applications of the Torrance Tests of Creative Thinking. Paper to XVIIth International Congress of App. Psychol. Belgium. July 1971.

TORRANCE, E.P. (1972) Can we teach children to think creatively? *J.Creat.Behav.*, 6, 114-143.

TORRANCE, E.P. (1974) *Torrance Tests of Creative Thinking: Norms-Technical Manual; Directions Manuals and Scoring Guides for Verbal Tests A and B, and Figural Tests A and B.* Lexington, Mass.: Personnel Press.

TORRANCE, E.P., and ALIOTTI, N.C. (1969) Sex differences in levels of performance and test-retest reliability on the Torrance Tests of Creative Thinking Ability. *J.Creat. Behav.*, 3, 52-57.

TORRANCE, E.P., and DAUW, D.C. (1965) Mental health problems of three groups of highly creative high school seniors. *Gifted Child Quarterly*, 9, 123-127.

TORRANCE, E.P., and KHATENA, J. (1970) *Technical-norms manual for 'What Kind of Person Are You?'* Athens, G.A.: Univ. of Georgia Studies of Creative Behaviour.

TORRANCE, E.P., TAN, C.A., and ALLMAN, T. (1970) Verbal originality and teacher behaviour: a predictive validity study. *J.Teach.Ed.*, 21, 335-341.

TREFFINGER, D.J., SPEEDIE, S.M., and BRUNNER, W.D. (1974). Improving children's creative problem-solving ability: The Purdoe Creativity Project. *J.Creat.Behav.*, 8, 20-30.

VAN MONDFRANS *et al* (1971) The effects of instructions and response time on divergent thinking test scores. *Psychol.in Schools*, 8, 64-71.

VAUGHAN, M., and MYERS, R.E., An examination of musical process as related to creative thinking. *J.Res.Music.Ed.*, 19, 337-341.

VERNON, P.E. (1961) *The Structure of Human Abilities* (2nd edn.) London: Methuen.

VERNON, P.E. (1964) Creativity and intelligence. *Educ.Res.*, 6, 163-169.

VERNON, P.E. (1967) Psychological studies of creativity. *J.Clin.Phys.Psychol.*, 8, 153-164.

VERNON, P.E. (1969) *Intelligence and Cultural Environment.* London: Methuen.

VERNON, P.E. (1971) Effects of administration and scoring on divergent thinking tests. *Br.J.educ.Psychol.*, 41, 245-257.

VERNON, P.E. (1972) The validity of divergent thinking tests. *Alberta J.Educ.Res.*, XVIII, 4, 249-258.

VERNON, P.E. (1973)(Ed.) *Creativity.* Harmondsworth, Middlesex: Penguin.

VERNON, P.E., ADAMSON, G., and VERNON, D.F. (1977) *The Psychology and Education of Gifted Children.* London: Methuen.

VINACKRE, W.E. (1952) *The Psychology of Thinking.* N.Y.: McGraw-Hill.

WALLACH, M.A. (1970) Creativity. In Mussen, P.H. (Ed.) *Carmichael's Manual of Child Psychology*. Vol.1. N.Y.: Wiley.

WALLACH, M.A. (1972) Torrance tests of creative thinking. Review. In Buros, O.K. (Ed.) *7th Mental Measurements Yearbook.* N.Y.: Gryphon Press.

WALLACH, M.A., and KOGAN, N. (1965) *Modes of Thinking in Young Children.* N.Y.: Holt, Rinehart and Winston.

WALLACH, M.A., and WING, C.W. (1969) *The Talented Student.* N.Y.: Holt, Rinehart and Winston.

WALLAS, G. (1926) *The Art of Thought.* N.Y.: Harcourt Brace.

WARBURTON, F.W. (1970) The British intelligence scale. In Dockrell, W.B. (Ed.) *On Intelligence.* London: Methuen.

WARD, James (1967) An oblique factorisation of Wallach and Kogan's 'creativity' correlations. *Br.J.educ.Psychol.*, 37, 380-382.

WARD, W.C., KOGAN, N., and PANKOVE, E. (1970) Motivation and capacity in children's creativity. Research Bulletin. RB-70-39. Princeton, N.J.: Educational Testing Service.

WECHSLER, D. (1949) *Wechsler Intelligence Scale for Children.* N.Y.: Psychol.Corporation.

WECHSLER, D. (1958) *The Measurement and Appraisal of Adult Intelligence.* London: Balliere, Tindall and Cox. Extract in *Selected Papers of David Wechsler* (1974). London: Academic Press.

WERSBERG, P.S. and SPRINGER, K.J. (1961) Environmental factors in creative production. *Arch.Gen.Psychiatry*, 5, 554-564.

WELSH, G.S. (1973) Perspectives in the study of creativity. *J.Creat.Behav.*, 7, 231-246.

WERTHEIMER, M. (1961, 1st edn. 1949) *Productive Thinking.* London: Tavistock.

WESMAN, A.G. (1952) Reliability and confidence. In Mehrens, W.A. (Ed.)(1976) *Readings in Measurement and Evaluation in Psychology and Education.* N.Y.: Holt, Rinehart and Winston.

WHITE, J.P. (1968) Creativity and education: a philosophical analysis. *Br.J.Educ.Studies*, 16, 2, 123-137.

WHITE, R.W. (1961) Motivation reconsidered: the concept of competence. *Psychol.Rev.*, 66, 297-333.

WISEMAN, S. (1949) The marking of English compositions in grammar school selection. *Br.J.educ.Psychol.*, 19,(3), 1949.

WISEMAN, S. (1956) The use of essays in selection at 11+. Reliability and validity. *Br.J.educ.Psychol.*, 26,(3), 1956.

WISEMAN, S. (1966) *Correlation Methods. Statistical Guides in Educational Research, No.1.* Manchester: Manchester Univ.Press.

WISEMAN, S. (Ed.)(1967) *Intelligence and Ability: Selected Readings.* England: Penguin.

WISEMAN, S., and START, K.B. (1965) A follow-up of teachers five years after completing their training. *Br.J.educ. Psychol.*, 35, 342-361.

WITKIN, H.A. (1965) Psychological differentiation. *J.Abnorm. Psychol.*, 70, 317-36.

WITT, G. (1971) The life enrichment activity program. *J.Res. Devel.Educ.*, 4, 14-22.

WODTKE, K. (1963) A study of the reliability and validity of creativity tests at the elementary school level. Unpublished Ph.D. thesis. Univ.of Utah. University Microfilms. High Wycombe, England.

WOOTON, W. (1965) *S.M.S.G. The Making of a Curriculum.* New Haven: Yale Univ.Press.

WYLIE, P.B. (1976) Effects of course grouping and skewed marginal distributions on the Pearson product-moment correlation coefficient. 36, 1-7.

YAMAMOTO, K. (1962) A study of the relationships between creative thinking abilities of fifth grade teachers and academic attainment. Doctoral dissertation. Minneapolis: Univ.of Minnesota.

YAMAMOTO, K. (1965) Creativity: A blind man's report on the elephant. *J.Counselling.Psychol.*, 12(4), p.428.

YAMAMOTO, K. (1965b) *Experimental Scoring Manuals for the Minnesota Tests of Creative Thinking.* Kent, Ohio: Bureau of Ed.Res. Kent State Univ.

YAMAMOTO, K. (1965c) Effects of restriction of range and test unreliability on correlation between measures of intelligence and creative thinking. *Br.J.educ.Psychol.*, 35, 300-305.

YAMAMOTO, K., and FRENGEL, B.A. (1966) An exploratory component analysis of the Minnesota tests of creative thinking. *Calif.J.Ed.Res.*, 17, 220-229.

YAMANE (1967) *Statistics: An Introductory Analysis.* London: Harper and Row.

ZEGAS, J. (1976) A validation study of tests from the divergent production plane of the Guilford structure-of-intellect model. *J.Creat.Behav.*, 10, 170-177 and 188.

APPENDIX

Long-Term Stability

Means and Standard Deviations of Basic Variables and Intercorrelations between all Variables Including Combined Scores

		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Mean	S.d.
(Flu.	1	86	64	91	47	48	38	48	61	51	59	31	31	32	48	42	48	33	42	50	60	40	23	04	19.27	7.60
Circles 1(Flex.	2		79	96	41	44	40	46	64	58	64	37	38	39	57	49	56	38	43	43	68	46	32	07	13.91	5.56
(Orig.	3			88	32	36	41	40	47	41	46	29	33	32	43	42	45	28	29	31	52	35	17	-01	6.26	6.48
Combined Circles 1	4				43	46	43	49	62	54	61	35	37	37	54	49	54	36	42	41	65	44	26	04		
(Flu.	5					92	67	95	38	29	35	53	47	52	30	36	35	50	50	53	40	48	13	06	23.81	6.74
Circles 2(Flex.	6						65	94	41	27	36	51	46	51	31	38	37	52	47	53	41	58	15	11	19.32	4.93
(Orig.	7							85	41	36	40	49	55	54	33	38	38	49	47	51	44	58	26	-10	14.72	11.12
Combined Circles 2	8								44	34	41	56	54	57	35	41	40	55	53	57	45	64	20	03		
Uses 1(Flu.	9									81	95	46	43	46	63	49	60	42	41	44	87	50	46	21	23.26	8.65
(Orig.	10										95	48	46	49	52	42	50	37	40	41	82	50	53	11	22.10	15.79
Combined Uses 1	11											50	47	50	61	48	58	41	43	44	89	53	52	17		
Uses 2(Flu.	12												86	97	41	36	40	59	58	62	51	89	34	27	31.19	10.66
(Orig.	13													97	36	34	37	53	53	56	47	85	33	18	30.16	19.06
Combined Uses 2	14														40	36	40	58	58	61	51	90	35	23		
Conse- (Flu.	15															78	94	44	41	45	87	47	47	20	9.78	3.45
quences 1(Orig.	16																94	42	40	43	80	44	37	11	7.14	3.98
Combined Conseq.1	17																	45	43	47	89	48	44	17		
Conse- (Flu.	18																		78	94	49	84	38	29	18.74	6.80
quences 2(Orig.	19																			94	48	84	43	13	9.97	5.12
Combined Conseq.2	20																				51	90	42	22		
Combined Verbal 1	21																					57	54	19		
Combined Verbal 2	22																						43	25		
I.Q.	23																							10	101.51	12.00
Sex(boy 0,girl 1)	24																									

(decimal points omitted in the table)

Main 11-year old Sample - Means and Standard Deviations

		Girls (n = 84)		Boys (n = 77)		Whole Population (n = 161)	
		Mean	s.d.	Mean	s.d.	Mean	s.d.
1.	I.Q.1	97.55	24.20	97.39	13.81	97.47	13.97
2.	I.Q.2	98.01	13.19	98.99	13.03	98.48	13.08
3.	Attitude to School	7.40	2.60	6.75	3.02	7.09	2.82
4.	Importance of Doing Well	6.67	1.99	6.94	1.76	6.80	1.88
5.	Conforming	2.50	1.16	2.70	1.11	2.60	1.14
6.	Relationship with Teacher	2.70	1.65	2.03	1.69	2.37	1.70
7.	Lack of Anxiety	2.68	1.09	2.99	1.13	2.83	1.12
8.	Academic Self-Image	8.81	2.74	10.04	3.17	9.40	3.01
9.	Self-Concept	55.98	7.96	57.13	6.44	56.53	7.28
10.	Extraversion	18.58	2.90	17.99	3.80	18.30	3.37
11.	Neuroticism	14.82	4.61	13.75	5.18	14.31	4.91
12.	Lie Scale	3.42	2.05	3.44	2.01	3.43	2.02
13.	Leisure Interests - Torrance	40.44	16.41	45.42	15.52	42.82	16.13
14.	Creative Motivation - Golann	20.90	4.91	20.94	4.21	20.92	4.57
15.	Circles (Fluency)	20.54	6.07	18.86	6.12	19.73	6.13
16.	Circles (Flexibility)	15.20	4.76	14.05	4.38	14.65	4.60
17.	Circles (Originality)	11.93	9.07	13.81	9.74	12.83	9.42
18.	Uses (Fluency)	11.81	4.69	10.16	3.86	11.02	4.38
19.	Uses (Originality)	9.58	7.06	7.17	5.75	8.43	6.56
20.	Consequences (Fluency)	9.83	3.72	7.94	3.51	8.93	3.73
21.	Consequences (Originality)	6.96	6.08	5.91	5.62	6.46	5.87
22.	Board Game	1.11	1.43	2.26	1.77	1.66	1.70
23.	Interests (overall)	20.96	4.72	17.74	4.90	19.42	5.06
24.	Imaginative Interests	8.30	2.35	6.00	2.62	7.20	2.73
25.	Logical/analytical Interests	4.02	1.73	5.60	1.70	4.78	1.88
26.	Creative Behaviour - Teacher Ratings	8.88	4.61	7.86	5.09	8.39	4.85
27.	English	3.26	0.88	2.84	1.04	3.06	0.98
28.	Maths	3.05	0.90	2.96	1.04	3.01	0.97
29.	Shy	2.87	1.02	2.73	1.29	2.80	1.56
30.	Lively	3.14	1.00	3.12	1.20	3.13	1.10
31.	Pleasurability	4.02	0.98	3.61	1.03	3.83	1.02

Main 11-year old Sample (Girls n=84)

Correlations between Variables 1 to 31

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	87	31	40	-11	31	11	39	18	27	02	-18	18	33	33	34	35	23	27	29	38	29	24	22	21	70	79	76	-02	15	36
2		35	35	-10	25	20	51	26	28	00	-23	12	34	22	22	28	14	20	22	27	33	14	15	13	62	78	75	-03	19	35
3			44	32	51	20	45	49	26	-13	14	22	38	08	14	05	15	24	23	12	28	32	44	19	19	45	39	-06	01	41
4				23	40	-16	51	39	12	07	02	09	15	19	23	10	20	14	27	11	10	25	21	13	38	51	46	10	-10	39
5					34	-06	13	09	-23	-15	43	-09	19	-11	04	-09	05	00	-13	-23	16	04	11	08	-15	-02	-02	12	-21	14
6						06	44	27	11	-12	31	-05	22	-02	13	05	09	-01	01	07	14	26	33	03	21	38	38	16	-16	52
7							08	08	28	-33	-05	09	24	12	04	05	-09	-04	03	00	12	05	12	06	07	08	14	-14	09	-02
8								47	33	-14	08	20	23	17	17	03	13	14	27	12	23	21	16	16	40	49	45	-01	09	40
9									41	-13	16	27	22	02	08	05	10	12	21	20	-13	25	29	09	19	28	19	-08	16	20
10										-13	-23	37	26	20	15	21	11	15	21	24	05	14	17	07	30	28	22	-30	34	10
11											-22	14	-12	15	21	02	28	22	28	22	-14	-08	-11	-05	-09	03	08	28	-11	04
12												-02	00	04	05	-02	04	-07	00	-12	-04	03	06	-01	-27	-14	-17	25	-26	10
13													29	41	33	18	37	27	41	24	-01	52	51	34	25	20	03	-21	18	10
14														21	14	15	25	32	27	26	24	35	51	21	35	36	24	-23	14	15
15															79	63	61	56	54	41	25	08	10	18	26	29	23	-08	11	09
16																68	62	43	45	42	29	12	14	16	17	28	29	06	07	23
17																	43	44	33	50	23	03	08	18	21	28	20	-07	13	14
18																		66	59	44	13	24	25	16	21	26	22	04	08	24
19																			57	53	22	15	22	20	32	32	26	-06	13	16
20																				41	-05	24	23	20	21	30	25	-07	14	20
21																					17	16	14	24	34	30	26	-11	23	15
22																						05	00	24	14	23	31	02	03	30
23																							74	69	18	19	07	09	-10	25
24																								26	28	29	15	-08	02	29
25																									14	05	-03	01	-02	16
26																										67	61	-36	40	28
27																											80	-02	12	44
28																												09	11	47
29																													-73	40
30																														-24
31																														

Decimal points omitted in the table.

(key to titles of variables on previous page with Mean Scores)

Main 11-year old Sample (Boys n=77)

Correlations between Variables 1 to 31

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	83	12	33	-04	11	35	48	04	12	-08	-17	09	17	24	09	40	15	37	30	38	38	05	02	20	69	79	72	-23	38	21
2		23	37	13	20	42	55	18	11	-10	-13	06	20	24	07	34	26	33	31	33	35	07	04	21	66	70	73	-16	28	34
3			46	41	47	35	34	46	16	-15	17	03	20	19	13	16	08	-12	14	10	15	25	28	07	15	16	15	-17	17	27
4				46	34	17	26	11	-14	13	01	-20	00	-08	-23	07	05	15	13	08	16	07	06	13	20	23	31	-03	03	26
5					34	11	09	02	-20	14	18	-25	01	-13	-10	-07	-11	-13	-14	-12	19	-01	08	-11	-05	-01	06	11	-12	31
6						36	36	42	22	-10	22	-15	04	-15	-19	-15	-26	-23	-12	-13	00	14	18	16	20	06	16	-01	-04	36
7							39	44	39	-32	04	17	20	17	07	11	16	-01	22	09	30	18	11	16	29	26	38	-10	12	20
8								49	30	-16	10	13	06	19	06	17	10	04	30	28	23	10	00	33	49	51	62	-19	32	38
9									53	-15	26	08	17	05	-03	04	-02	-14	11	12	03	28	27	15	22	10	18	-25	21	28
10										-18	01	15	22	-09	-16	00	-19	-08	04	08	-10	15	17	07	26	20	27	-25	25	16
11											-20	09	07	01	04	02	00	11	02	-06	-06	-11	05	-23	-03	-15	-18	-10	13	02
12												-12	00	-16	-15	-11	-06	-31	-15	-13	-07	22	15	20	-16	-14	-10	04	-12	21
13													18	31	30	24	29	05	32	12	21	22	14	15	17	13	10	-25	25	-10
14														19	11	20	18	18	14	22	11	03	07	12	28	18	10	-07	-02	33
15															87	70	53	34	68	43	28	13	09	20	24	26	13	-28	31	-05
16																63	47	26	62	38	22	01	00	-03	09	13	00	-16	25	-14
17																	28	33	58	44	35	17	15	14	41	37	22	-39	30	09
18																		56	47	18	39	15	-02	21	11	13	10	00	07	-04
19																			36	33	27	-01	00	05	26	29	25	-11	15	05
20																				58	22	08	05	12	28	40	24	-32	31	03
21																					18	09	09	09	38	54	39	-25	24	11
22																						17	03	21	43	28	22	-30	33	11
23																							81	52	09	10	11	-27	03	23
24																								17	03	04	03	-35	14	24
25																									25	19	33	-04	-09	29
26																										70	60	-49	56	35
27																											78	-21	33	37
28																												-15	30	38
29																													-71	-04
30																														04
31																														

Decimal points omitted in the table.

(key to titles of variables on previous page with Mean Scores)

Main 11-year old Sample (Whole Population n=161)

Correlations between Variables 1 to 31

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	85	21	37	-08	21	22	42	12	18	-03	-18	13	26	29	23	37	19	31	28	38	31	15	11	18	69	77	74	-13	27	29
2		28	36	01	21	31	53	23	18	-06	-18	09	28	22	14	31	18	24	24	29	33	09	07	17	63	71	73	-10	24	33
3			43	35	50	26	35	45	21	-13	15	10	24	15	15	09	14	09	21	12	15	30	37	07	18	31	26	-11	10	35
4				34	35	00	39	28	-02	09	01	-02	09	06	02	09	12	13	18	09	15	14	10	15	28	35	38	03	-03	31
5					31	03	13	07	-22	-01	31	-15	11	-13	-04	-07	-03	-07	-15	-18	19	-01	05	03	-11	-03	01	11	-16	20
6						18	34	31	18	-09	26	-13	13	-05	01	-07	-02	-06	00	00	-01	25	31	00	17	25	27	08	-09	46
7							27	24	32	-33	00	14	22	12	04	09	-01	-05	08	03	25	06	04	15	16	14	25	-12	10	01
8								48	28	-17	09	19	14	15	08	12	07	05	21	17	28	08	-02	30	42	43	52	-12	21	33
9									45	-14	20	20	20	02	03	05	04	01	14	16	-03	22	22	14	19	17	18	-16	18	21
10										-15	-10	24	23	05	00	08	-02	05	14	16	-07	16	19	02	29	25	25	-26	29	15
11											-21	10	-03	09	14	01	17	18	18	09	-13	05	02	-17	-05	-04	-06	07	02	05
12												-06	00	-06	-04	-06	00	-17	-07	-13	-05	11	09	08	-22	-14	-13	14	-19	15
13													24	34	29	22	29	14	31	17	15	30	22	29	19	12	06	-23	21	-03
14														20	13	17	22	26	21	24	16	19	28	16	32	27	17	-15	06	22
15															83	64	58	47	62	42	20	14	14	07	26	30	18	-17	22	05
16																63	56	37	53	41	19	10	12	02	14	22	15	-04	16	08
17																	33	36	41	46	31	06	06	19	30	30	20	-25	22	09
18																		63	56	34	17	25	19	08	18	23	16	03	08	15
19																			50	46	16	13	18	04	31	33	25	-07	14	15
20																				49	-01	23	23	04	26	39	24	-18	22	16
21																					13	15	14	12	36	42	32	-17	23	14
22																						-01	-13	33	25	16	23	-18	19	11
23																							80	39	16	20	10	-08	-03	29
24																								00	18	23	10	-18	09	32
25																									13	02	12	-04	06	11
26																										69	61	-42	48	33
27																											78	-11	24	43
28																												-05	22	42
29																													-71	17
30																														-09
31																														

Decimal points omitted in the table.

(key to titles of variables on previous page with Mean Scores)

Principal-Factor Analysis with Varimax Rotation

Girls (n = 84)

	FACTORS 1	2	3	4	5	6	7	8	h^2
I.Q.1	85	25	09	-02	-12	04	13	08	82
I.Q.2	85	13	01	-05	-12	11	08	16	80
English	85	21	17	-02	02	16	-07	-01	82
Maths	87	17	01	06	01	09	-12	02	82
Attitude to School	36	08	30	00	42	37	05	14	57
Importance of Doing Well	47	11	05	09	27	39	10	-28	55
Conforming	-03	-07	05	06	76	-02	05	-06	59
Relationship with Teacher	39	-04	17	22	48	25	-04	08	54
Lack of Anxiety	08	01	11	-06	-05	06	-03	59	37
Academic Self-Image	47	07	-01	-02	23	54	14	05	60
Self-Concept	15	04	18	-08	13	72	-01	04	60
Circles (Flu.	10	86	-02	-03	-03	04	05	12	77
Circles (Flex.	13	82	-04	11	08	05	06	09	73
Circles (Orig.	16	67	-07	-03	-03	-05	08	18	55
Uses (Flu.	08	75	23	01	09	03	-01	-23	69
Uses (Orig.	17	67	19	-13	05	-02	02	-17	56
Consequences (Flu.	12	61	23	-04	-09	27	-01	-19	56
Consequences (Orig.	23	54	09	-12	-15	07	10	-05	41
Board Game	34	22	-12	00	26	-21	27	26	43
Interests (overall)	09	05	74	16	00	18	57	-01	94
Imaginative Interests	14	10	87	00	14	11	06	09	84
Logical/analytical Interests	04	17	24	02	02	04	83	02	78
Creative Behaviour Teacher Ratings	69	16	17	-37	-11	09	06	-03	70
Leisure Interests Torrance	-03	38	46	-12	-17	30	19	03	54
Creative Motivation Golann	28	20	46	-21	20	03	05	19	46
Extraversion	19	17	12	-23	-23	47	00	32	51
Neuroticism	00	27	-02	22	-27	-06	-11	-53	49
Lie Scale	-23	03	01	24	52	13	-04	06	40
Shy	04	-02	-08	93	06	-06	02	-15	90
Lively	10	12	-02	-73	-19	12	-03	08	62
Pleasurability	47	13	14	38	24	17	08	-02	50
Percentage of Common Variance	25.0	21.4	11.5	10.5	9.6	9.1	6.5	6.2	
Percentage of Total Variance	15.7	13.4	7.2	6.6	6.0	5.7	4.1	3.9	

Common Variance extracted = 62.7%

(decimal points are omitted in the table)

Boys (n = 77)

	FACTORS 1	2	3	4	5	6	7	8	9	h^2
I.Q.1	86	12	05	07	00	14	16	-09	-03	82
I.Q.2	79	12	15	21	-01	05	23	-04	07	77
English	87	23	05	01	03	08	-06	01	11	83
Maths	85	05	18	05	03	04	02	08	04	77
Attitude to School	05	20	43	60	16	09	00	09	-01	63
Importance of Doing Well	32	-09	-01	65	08	-03	11	-10	00	57
Conforming	-03	-10	-05	7 6	-03	-03	00	05	07	60
Relationship with Teacher	10	-18	42	48	07	-01	-14	21	05	52
Lack of Anxiety	27	08	65	14	03	-01	26	05	-06	60
Academic Self-Image	53	13	40	18	-07	14	02	29	06	61
Self-Concept	04	04	70	14	15	16	-09	18	19	64
Circles (Flu.	09	90	03	-04	03	11	19	00	01	87
Circles (Flex.	-07	91	-03	-06	-08	07	12	00	-03	87
Circles (Orig.	26	67	-07	00	11	22	10	00	09	60
Uses (Flu.	06	44	-02	-06	04	-11	71	-04	10	73
Uses (Orig.	32	26	-20	-08	05	-03	40	-34	20	53
Consequences (Flu.	25	72	09	-03	04	10	13	-10	05	63
Consequences (Orig.	43	49	01	-08	08	04	-08	-09	12	47
Board Game	25	17	-03	16	00	32	52	13	-01	51
Interests (overall)	04	06	11	00	93	05	15	30	-01	99
Imaginative Interests	-04	06	12	13	85	16	-07	-01	09	80
Logical/analytical Interests	28	01	06	-08	31	-10	25	54	10	56
Creative Behaviour Teacher Ratings	69	10	10	-01	-04	48	12	07	24	81
Leisure Interests Torrance	01	27	15	-27	13	21	23	01	05	29
Creative Motivation Golann	11	14	15	-03	02	-02	13	-04	57	40
Extraversion	16	-12	68	-26	11	18	-19	-06	22	70
Neuroticism	-16	00	-31	14	00	21	-01	-31	23	33
Lie Scale	-17	-10	14	17	13	-06	-08	43	03	30
Shy	-13	-19	-10	04	-25	-76	01	06	-02	71
Lively	25	20	12	-03	-01	79	02	-12	-06	77
Pleasurability	33	-07	05	33	15	02	-11	30	57	67
Percentage of Common Variance	22.4	17.9	11.2	10.5	9.7	9.4	7.2	5.8	5.0	
Percentage of Total Variance	15.0	11.5	7.2	6.7	6.2	6.0	4.6	3.7	3.2	

Common Variance extracted = 64.1%

(decimal points are omitted in the table)

<u>Principal-Factor Analysis with Varimax Rotation</u>

Appendix

Inter Scorer Reliability

Marks awarded by Scorer 1 and Scorer 2

Users test (n=42)

Fluency		Originality	
Scorer 1	Scorer 2	Scorer 1	Scorer 2
30	31	19	15
14	14	20	19
6	6	3	6
13	13	17	18
6	5	1	2
7	7	6	8
44	43	41	26
15	15	11	15
36	36	20	24
14	14	19	20
8	8	8	4
23	23	24	21
20	19	29	35
27	27	33	28
7	7	4	7
7	7	5	4
8	8	7	8
16	16	14	19
10	10	11	10
25	25	40	45
2	1	0	0
24	23	18	16
25	25	34	31
12	11	3	5
11	11	7	9
8	5	3	8
25	18	24	17
20	20	15	17
9	9	2	2
14	14	15	18
13	12	15	18
21	19	19	17
20	20	20	26
7	3	0	1
11	11	12	10
11	12	21	22
10	10	9	12
11	11	5	8
3	2	0	1
7	6	6	12
15	15	12	15
19	17	30	47

Consequences test (n=41)

Fluency		Originality	
Scorer 1	Scorer 2	Scorer 1	Scorer 2
6	6	3	5
8	8	6	5
8	7	7	9
17	17	3	3
13	14	9	10
27	27	22	24
10	10	11	12
14	14	28	26
8	8	11	15
12	12	3	3
11	11	6	7
8	8	10	5
11	11	7	9
13	12	15	17
13	12	6	8
16	16	12	8
15	14	11	10
12	12	12	13
6	6	1	3
5	5	0	0
19	19	16	14
8	8	0	0
7	7	0	1
5	5	2	2
6	6	2	1
12	12	10	14
9	9	6	8
6	6	2	2
16	16	21	19
22	21	15	18
5	5	0	0
4	6	6	10
11	11	3	4
20	22	23	24
9	9	7	7
6	6	3	3
15	13	15	12
6	6	10	11
6	7	6	5
8	8	3	2
15	15	21	23

Circles test (n=42)

Fluency		Flexibility		Originality	
Scorer 1	Scorer 2	Scorer 1	Scorer 2	Scorer 1	Scorer 2
16	16	14	13	20	23
24	24	18	19	16	23
34	31	20	23	15	11
11	11	8	8	9	8
24	24	19	21	39	34
16	14	10	10	14	6
17	16	10	10	7	3
13	13	13	13	11	10
30	29	22	21	42	32
26	26	18	18	45	38
11	14	11	12	16	12
12	12	6	8	2	1
19	19	15	16	12	12
16	16	12	11	23	24
16	17	11	12	7	6
16	15	11	12	5	4
23	23	17	15	6	2
15	15	13	14	19	19
10	10	10	9	9	7
21	21	16	14	22	25
10	10	6	6	2	7
21	21	15	15	21	16
22	22	19	18	22	18
20	20	16	15	28	27
18	18	13	13	8	7
29	29	20	21	37	46
17	17	12	12	17	18
25	25	15	15	21	19
12	12	9	9	7	7
16	16	16	15	19	22
31	31	18	20	31	33
22	22	14	14	9	11
18	18	12	15	32	34
14	14	10	8	11	9
17	17	11	12	15	21
19	18	9	8	0	0
22	22	15	14	26	25
19	19	12	12	23	21
11	11	9	10	13	12
41	42	26	31	82	71
17	16	8	9	19	8
12	12	6	6	0	0

Flexibility Categories for use with the Circles Test and the Squares Test

1. Balls: football, golf ball, ball of wool, bolas etc
2. Balloons: air balloon, toy balloon, bubbles etc
3. Bats and racques: tennis racket etc
4. Books and stationery: book, comic, birthday card, exercise book, writing paper etc
5. Bottles and jars: jam jar, bottle, vase etc
6. Boxes: toy box, lunch box, cereal packet, parcel etc
7. Buildings: houses, block of flats, light-house, tent, igloo etc
8. Cages: animal cages, kennel, hutch, jail etc
9. Clothing: skirt, hat, shoe etc
10. Coins
11. Containers for carrying: bag, shopping basket, bucket, handbag etc
12. Other containers: bin, barrel etc
13. Crockery: tea-cup, plate,saucer etc
14. Cross-section/end of 'open' objects: tunnel, pipe, manhole etc
15. Cross-section/end of solid objects: tree, log, post, worm, bone etc
16. Dials: Clock, watch, compass etc
17. Electric light and fittings: light bulbs, headlights, plug, torch etc
18. Electrical appliances: cooker, washing machine, refrigerator etc
19. Faces (Human): pirate, king, queen, sad face etc
20. Faces (animal): cat, dog, rabbit, pig etc
21. Figures (human): fat man, baby, humpty dumpty etc
22. Figures (animal): cat, octopus, spider etc
23. Flowers: sunflower, poppy etc
24. Food: fried egg, cake, hot X bun, Christmas pudding etc
25. Fruits: apple, orange, bunch of grapes, berries etc
26. Furniture: table, chair, stool etc
27. Games: ludo, O's and X's, crossword puzzle etc
28. Geometrical shapes and patterns: circle, compass pattern, spiral, hexagon, square, triangle etc
29. Head of a nail: screw, pin etc
30. Heavenly bodies: star, sun, moon, planet, globe etc

31. Jewellry: ring, necklace, hair slide, crown etc
32. Land forms: volcano, crater, island etc
33. Optical instruments: binoculars, magnifying glass, spectacles, goggles etc
34. Lids: top of jam jar, tobacco tin etc
35. Linen and mats: table cloth, mat, towel etc
36. Microscopic views: bloodcell, germ, dust particle, amoeba etc
37. Musical instruments: bagpipes, banjo, trumpet etc
38. Parts of the body: eye, mouth, nose, hand etc
39. Parts of a building: door, window, chimney, turret etc
40. Pictures: pictures, photographs, paintings etc
41. Pots and pans: frying pan, saucepan, mixing bowls etc
42. Rotors: propellor, helicopter, fan, etc
43. Signs: (i) road signs, shop signs etc
44. Signs: (ii) emblems, badges etc
45. Sound and television: radio, television, microphone, record player etc
46. Sweets: toffee, gob-stopper, lollipop etc
47. Symbols: Letters of alphabet, numerals, musical notes etc
48. Target: dartboard, bull's eye etc

49-52. Transport and vehicles:

  49. Vehicles needing propulsion: pushchair, wheelchair shopping trolley, bicycle etc
  50. Land travel: car, lorry, train, tank etc
  51. Sea Travel: ship, submarine etc
  52. Air and Space travel: aeroplane, rocket, satelite, spaceship etc

53. Trees: fruit tree, Christmas tree etc
54. Water: pond, puddle, swimming pool etc
55. Weapons: gun, bomb, mine, gun, turret etc
56. Wheels: steering wheel, water wheel etc
57. Vegetables: beetroot, turnip etc

Frequency Distribution of Responses to the Circles Test

11 year old sample, n = 165

Allocation of Originality Marks

Frequency (percentage)	f≤1%	1%<f≤5%	5%<f≤10%	10%<f≤15%	f>15%
Frequency (number of responses)	1 (unique)	2-8	9-16	17-24	24
Originality Mark	5	3	2	1	0

NOTE:

Some of the common responses could be considered for originality marks if they have some special defining feature in relation to a circle that removes them from an unspecified category - e.g. Kojak (human face), crystal ball (ball), humpty dumpty (human figure) - see discussion in the text.

Response	Orig. Mark
Air balloon	1
Aerial view of Mexican, man on skiis	3
Ashtray	3
Acorn	3
Atom	3
Artist's pallette	3
Apple fritter	5
Amoeba	5
Amphitheatre	5
Ball	0
Balloon	0

Response	Orig. Mark
Bat/racquet	0
Bicycle	0
Button	0
Bubble	1
Bomb	2
Buckle	2
Bell (electric)	2
Bicycle bell	2
Bottle/jar (round section)	2
Banjo/guitar/ ukelayli	2

Response	Orig. Mark	Response	Orig. Mark
Bee	2	Bat (flying)	5
Butterfly	2	Berry	5
Badge	2	Buoy	5
Bell-push	2		
Bracelet	3	Baseball glove	5
Bag (Handbag)	3	Boxing glove	5
Bow	3	Boat engine	5
Bubble car	3	Bikini top	5
Battery	3	Bagpipes	5
Bin	3	Bloodcell	5
Barrel	3	Basket of air balloon	5
'Big wheel' at fun fair	3	Belt	5
Bud	3	Bolas	5
Bath plug	3	Batherscape	5
Bottle of perfume etc (circular elevation)	3	Coin	0
Biscuit	3	Crockery	0
Binoculars	3	Cross-section of tree trunk/log	0
Bead	3		
Back view of pig/ elephant	3	Compass	1
Back view of person bending	3	Cake	1
Belisha beacon	3	Car/lorry	1
Bird cage	3	Circle	2
Bee-hive	3	Cannon ball	2
Ball and chain	3	Cogwheel	2
Birthday card	3	Chair	2

Response	Orig. Mark	Response	Orig. Mark
Cotton reel	3	Cartoon hand	5
Cooker ring	3	Cylinder	5
50p coin	3	Chemical apparatus	5
Chimney pot	3	Dartboard	1
Cannon/attlery gun	3	Dumbells	2
Caterpilla	3	Doorknob	2
Cloud	3	Dog basket	3
Counter	3	Drum	3
Compass pattern	3	Diamond	3
Cartridge	3	Doughnut	3
Cabbage/cauliflower	3	Door number plate	3
Cuckoo clock hole	3	Dog biscuit	5
Cross-section of worm	3	Face of a Die	5
Crab	3	Flight end of a Dart	5
Cheese	3	Round Door (submarine)	5
Catseyes	3	Discus	5
Cobweb	3	Dust particle	5
Cherries	3	Diving bell	5
Conker	3	Demolition weight	5
Crystal ball	5	Eye	0
Cork	5	Egg/egg cup	2
Camera	5	Ear ring	3
Candle	5	Electric plug	3
Cloud over the moon	5	Egg hatching	5
Cave	5	Eye of a needle	5
Coal shute	5	Eye hole for shoe lace	5
Corn plaster	5	Elephants trunk (tip)	5
Caravan	5	Elephants footprint	5
Crown	5	Exhaust pipes (end)	5
Cigarette	5	Electron	5
Christmas pudding	5	Face (human or animal)	0
Citreon car badge	5	Flower	0
Candy floss	5	Fruit	0
Chinese puzzle	5	Fingernail	3

Response	Orig. Mark	Response	Orig. Mark
Filter funnel	3	Humpty dumpty	2
Flower vase	3	Helicopter	2
Frogs spawn	3	Headlights	2
Frog	3	Horseshoe	3
Fishing reel	3	Handgrenade	3
Fractions in a circle	3	Hoop	3
Frisbee	3	Hair bobbles	3
Foot	3	Hair slide	3
Fossil	5	Helmut (diver etc)	3
Finger print	5	Head (back of)	3
Fishing weight	5	Hand	3
Freckle	5	Horn	3
Fist	5	Hexagon	3
Flower bulb	5	Hot-X-bun	3
Flower bed	5	Head of a cane	5
Full stop	5	Heat lamp	5
Fly swatter	5	Hopscotch	5
Globe	0	Hubcap	5
Goldfish bowl	2	Helter-skelter	5
Gun barrel	3	Heart	5
Gun sights	3	Inkwell	1
Gyroscope	3	Ice cream	1
Ghost	3	Island	3
Gong	3	Ink blot	3
Gun turret	5	'Idea' circle	5
Gonk	5	Igloo	5
Goggles	5	Intestines	5
Goggles	5	Jar/bottle	2
Germ	5	Jellyfish	3
Gas ring	5	Jig-saw	3
Human or animal figure	0	Japanese flag	5
Hedgehog	1	Jewel case	5
Hole	1	Keyhole	3
Hat	2	Key ring	3

Response	Orig. Mark	Response	Orig. Mark
Key	3	Map	3
Kite	3	Mask	3
Kojak	3	Meteorite/comet/	
Keep (of a castle)	5	shooting star	3
Knot	5	Music note	3
Letter of alphabet/ numeral	0	Microphone	3
Light switch	0	Top of Monks head	3
Lollipop	0	Medal	3
Light bulb	1	Mobile	5
Lid	2	Megaphone	5
Ladybird	2	Magnet	5
Letter box	3	Meatball	5
Lens	3	Mop	5
Lampshade	3	Moon crater	5
Locket	3	Mace and chain	5
Leaf	3	Necklace	2
Lifebelt	3	Nest	2
Ladle	5	Nuts (screw)	3
Loaf of bread	5	Name disc	3
Link in a chain	5	Net	3
Lantern	5	Nostril/nose	3
Lorry (cement mixing)	5	Number on racing car	5
Leg of meat	5	Nut (fruit)	5
Ladies fan	5	Octopus	3
Ludo	5	Olympic sign	3
Mouth	0	Pots and pans	0
Magnifying glass	1	End of Pencil	1
Mirror	1	Porthole	1
Marble	2	Polomint	2
Monocle	3	Pencil sharpener	2
Manhole cover	3	Pinhead	2
Mine	3	Picture/painting	2
Mat	3	Pond/pool	2

Response	Orig. Mark	Response	Orig. Mark
Propellors	2	Road sign	0
Pig's nose	3	Record	2
Pipe (end of)	3	Rubber	2
Postmark	3	Roundel (RAF)	2
Punchball	3	Red Cross sign	3
Protractor	3	Rocket	3
Puffer fish	3	Rabbit hole	3
Pawn	3	Rope (end of)	3
Periscope	3	Radar	3
Playground	3	Ring (rubber)	3
Paddlesteamer	3	Roll of selotape/ wallpaper etc	3
Pebble	3		
Pendulum	3	Rock	5
Penny farthing	3	Ringo,	5
Purse	3	Reinforcement in pillar	5
Pram	3	Record player	5
Pea/bean	3	Raindrop	5
Peacock's tail	3	Railway turntable	5
Pepper pot	3	Snail	0
Plughole	3	Spectacles	0
Pineapple slice	3	Spider	0
Pipe (smokers)	3	Sun/moon/planet	0
Paper weight	3	Screw/nail head	1
Pulley	5	Snowman	2
Potato pattern	5	Sweet	2
Parachute	5	Spiral	2
Plug (for aerial)	5	Smilee	2
Pearl	5	Stool	2
Press-stud	5	Shell	2
Poached egg holder	5	Spaceman	3
Pin and string pattern	5	Snake	3
Pineapple slice	3	Switch (round)	3
Radio	0	Sea urchin	3
Ring (jewellery)	0	Shield	3

Response	Orig. Mark	Response	Orig. Mark
Ship's wheel	3	Tree	2
Starfish	3	Tortoise	2
Scissors	3	Tap	2
Spaceship	3	Tap washer	3
Speedometer	3	Torch	3
Seed	3	Telephone receiver	3
Spinning wheel	3	Tube	3
Spring	3	Tie	3
Soap	3	Toilet roll	3
Snowball	5	Toilet seat	3
Skull	5	Turnip/beetroot	3
Safe	5	Tent	3
Seal balancing a ball	5	Turret/tower	3
Skipping rope handle	5	Tunnel	3
Spray end of hose	5	Trophy/cup	3
Shark's mouth	5	Tray	3
Spinner	5	Tablet/aspirin	3
Shuttlecock	5	Tape measure	3
Stamp	5	Train	3
Sperm	5	Trundle wheel	3
Spurs	5	Top (spinning)	3
Sail	5	Teapot stand	5
Submarine	5	Tea strainer	5
Steamroller	5	Torpedo	5
Swimming pool	5	Trademark	5
Sack (of gold)	5	Trumpet	5
Shopping trolley	5	Ticket	5
Sausage roll	5	Thermometer	5
Soldier's water-bottle	5	Trout hatchery	5
Television	0	Toadstool	5
Traffic lights	1	Tambourine	5
Target	1	Tractor	5
Table	1	Tread wheel	5
Telephone (dial)	2	Umbrella/sunshade	3

Response	Orig. Mark	Response	Orig. Mark
U.F.O./flying saucer	3		
Unicycle	5		
Underground sign	5		
Volcano	3		
Vacuum cleaner	5		
Watch/clock	0		
Wheel	0		
Window	1		
Wool (ball of)	2		
Wash basin	3		
Weight	3		
Womble	3		
Wig	3		
Water wheel	3		
Well	3		
Windmill	3		
Washing machine (door)	3		
Weighing scales	3		
'Wood' for a game of bowls	5		
Wheelchair	5		
Wreath	5		
Wheelbarrow	5		
Windsock	5		
Whirlpool	5		
Yoyo	3		

Frequency Distribution of Responses to the Squares Test

11 year old sample, n = 65

Allocation of Originality Marks

Frequency (percentage)	f≤2%	2%<f≤5%	5%<f≤10%	10%<f≤15%	f>15%
Frequency (number of responses)	1 (unique)	2,3	4,5,6	7,8,9	10
Originality Mark	5	3	2	1	0

NOTE:

Some of the common responses could be considered for originality marks if they have some special defining feature in relation to a square that removes them from an unspecified category - e.g. Jack-in-the-box (box), block of flats (house), magic squares (square) - see discussion in the text.

Response	Orig. Mark
Abacus	5
Animal face	0
Animal body	0
Bat	3
Bale of hay	5
Bread and butter slice	5
Breadboard	5
Brick	1
BOOKS and STATIONARY	
Book (reading)	0
Book (exercise)	0
Birthday card/ Christmas card	2
Comic/newspaper	1
Folder	5
Printed paper/ Test paper/ Timetable	2
Envelope	0
Writing paper	0
Notice	2
Calendar	2
Blotting paper holder	5
Blackboard	0
Battery	5
Badge	3

Orig. Mark	Response	Orig. Mark	Response
Bow/bowtie	2	Teapot/kettle	1
Button	2	Tin can (round)	2
Buckle	3	Tin biscuits/cake tin	2
Brush	5		
Bolt (screw)	5	Crossword puzzle	2
Bonnet of car	5	Chip (end view)	5
Bird table	5	Cushion	3
Bar chocolate	2	CLOTHES	
BOXES		Shirt/shorts	2
Box (no labels)	0	Hat/tophat/soldiers hat	1
Box (carton for food etc)	0	Boot	3
Letter box	3	Door	0
Jack-in-the-box	2	Door knob/handle	2
Toy box/work box (with hinged lid)	2	Dice	2
		Doorbell	3
Lunchbox	3	Draughts board/ chess board	3
Box matches/cigatettes	3		
Box for Xmas tree	5	Drain cover	3
Bin (wastebin/dustbin)	2	Electric plug	0
Jute box	5	Electric light switch	0
Parcel	0	Field	3
Balance	3	Fish tank	1
Cage	2	Frame (e.g. around dartboard)	2
Camera	2		
Cake	3	Flag	0
Chimney	2	Fireplace/electric fire	3
Carpet/mat	2		
Cheese	3	Float	5
Control panel	3	Finger nail	5
CONTAINERS		Fractions of a square/ triangle	2
Cup/glass	1		
Water tank/petrol tank	2	FURNITURE	
Pan/pail/bucket	2	Chair/stool	0
Jars	2	Table	0
Dish	3	Cupboard/wardrobe	0

Orig. Mark	Response
Chest of drawers	2
Bed	1
Bookcase	2
Baby cot	3
Gate	0
Gravestone	1
Geometric pattern	2
HOUSES	
House	0
Barn/shed/garage	2
Animal houses, kennel/ Hutch	3
Roof	3
Block of flats/tower block/church/castle/ inside a room/ classroom	3
Handkerchief	3
Hopscotch square	5
Headlamp	5
Human face/robot face	0
Humanbody	0
Icecream cone/lollipop	3
Jigsaw	5
Keyring	3
Kite	2
Kitchen appliance: Fridge/cooker etc.	1
Lift	5
Lampshade	2
Label/price tag	2
Lighter	5
Loaf of bread	5
Letter/numeral O/P etc	3
Map	0
Mirror	3
Mobile (of squares)	5

Orig. Mark	Response
Net	3
O's and X's	3
Picture/Painting	0
Photograph	1
Pocket	2
Pencil Sharpener	3
Palette for Paints	3
Packet of Pencils/ pencil case	2
Pond	3
Piano	3
£1 note	3
Playground	5
Pencil	2
Pyramid	5
Postage Stamp	2
Radio	0
Rocket	2
Radiator	3
Roadsign	3
Record Player/Tape Recorder	1
Raft	5
Record (in jacket)	3
Rubber	0
Ruler	5
Square (magic square)	5
Sack	5
Sail	5
Safe	3
Swimming Pool	2
Suitcase	3
Sink	3
Shield	3
Screen (for films)	3
Scent Bottle (Channel No. 5)	5

Orig. Mark	Response
Ship (funnel)	3
Stand (for pencil)	3
Stage	5
Saddle	5
Steps/stair	3
Sewing Machine	5
Sausage on Tomorrow's World	5
Sports Pitch (football/ netball)	
Spiral	2
Spectacles	0
Sweet/toffee	0
Tape Measure	5
Tartan check	5
Tennis Racket (in press)	3
Teeth	5
Telephone	3
Television	0
Trampoline	3
Tile	2
Tree	2
Traffic Lights	3
VEHICLES	
Car - (given)	0
Lorry/Van/Bus	0
Trailer/Horse Box/ Caravan	3
Helicopter	3
Crane	3
Tank	2
Pram/Trolley	3
Train	3
Motor Bike (tyre from front)	5

Orig. Mark	Response
Ventilator	3
Wall	0
Washing Line	5
Washing Machine	2
Wallet	3
Watch/Clock	0
Weight/dumbels	3
Wood - plank/piece of	2
Well	5
Window (given)	0
Jail Window	2
Window Blind	2
Pane of Glass	3

Frequency Distribution of Responses to the Uses Test
(Uses for a Spoon)

Main 11-year old population (25% sample, n=42)

Allocation of Originality Marks

Frequency (percentage)	f≤3%	3%<f≤5%	5%<f≤10%	10%<f≤15%	f >15%
Frequency (number of responses)	1 (unique)	2	3,4	5,6	7
Originality Mark	5	3	2	1	0

Response	Orig. Mark
Armour	5
Breaking boiled egg	1
Break/smash things	1
Baby to bite on (chew)	3
Bend it	1
Bend like Uri Geller	3
Bend into circle for large earing or bracelet	5
Bird scarer	5
Bat	3
Back scratcher	3
Cooking/eating mixing, stirring, serving, mashing things up	1
Cake tin for baking	5
Clean up mess, scoop up spilt things	2
Catch tadpoles, small fish from pond	2
Ca rry/pick up small insects etc.	2
in Chemistry set (c.f. mixing paints etc.)	3
Collect (old spoons)	5
Dig with it	1
Dig weeds up, plant things with it, use as trowel in garden	2
Decoration make patterns with it paint different colours and hang it up draw on it, etc.	1
Draw round it for shapes, make shapes	1
Doctor or dentist use to see down throat/mouth	2
Drumstick	1
Eating/drinking, soup, medicine, etc.	1
Electricity conductor	2
Frying pan	5
Feed baby	2
Feed small animals	3
Float it (as boat)	3
Flick things (catapult)	1
Flick paint and make a modern picture	5

Response	Orig. Mark	Response	Orig. Mark
Get stuff out of tin/jar (e.g. dog food)	1	Nut cracker	3
Get clay out of bin	3	Play - see Games	1
Games		Pendulum	3
play with it, use it for tricks	1	Paddles	5
egg and spoon race	1	Play harp/guitar with	3
Give as a 'free spoon' for buying a product	5	Present (give as one)	2
Hitting with (on head, knuckles etc.), fight with	1	Paper weight/weight	3
Hammer	2	Plastering	5
Knife (sharpen it)	5	Put down your back to cool off	5
Juggling, balancing, spinning by hitting end	2	Prop something up, (e.g. plants)	2
Lever (for tyres)	3	Putting powder paints in trays/palettes and mixing paint	2
to get tin lids off	2	Pouring into/filling a jar	2
to open windows/ doors	2	Picking things up (food etc.)	1
Lift out boiled egg, hold something hot	3	Reflexes tester	3
Measuring (table-spoonful etc.)	1	Reflect light/make it shine in the sun	2
Making music (playing the spoons)	1	Rolling pin	1
Making xylophone from spoons or bottles and hitting	2	Spread/smooth some-thing out with it	3
Mirror	1	Scoop up mess - see Clean	2
Make noise (by banging it)	1	Swat flies, squash insects etc.	2
Mould for lollipop, etc.	2	Shoe horn	2
Make a hole in things	2	Scraping things	1
Make a doll	1	Stick for homemade lollies	5
Melt it down for metal	1	Spade - see Dig	1
Make a modern metal model/sculpture	3	Shovel for making sand castles, mud pies etc.	2
Make a blindfold/pair of spectacles	3	Stop boiling water cracking a glass	5
Marking something on the ground/marking patterns in plaster	3	Signal with it in the sun	5
		Screwdriver	3

Response	Orig. Mark
Specific spoons (e.g. egg, desert, teaspoon etc.) c.f. Eating/drinking	1
Shaping clay	3
Tap dancing (fix to shoes)	5
Test your strength (by bending)	2
Throw it	1
Wooden spoon (prize)	5
Wedge	1
Weight/paper weight	3
Wickets for cricket	5

Frequency Distribution of Responses to the Consequences Test

(Consequences of having no hair)

Main 11-year-old population (25% sample, n=42).

Allocation of Originality Marks

Frequency of a response in the population tested						
	Percentage cut-off points	$f \leq 3\%$	$3\% < f \leq 5\%$	$5\% < f \leq 10\%$	$10\% < f \leq 15\%$	$f > 15\%$
	Number range	1 (unique)	2	3,4	5,6	7
Originality Mark		5	3	2	1	0

Response	Orig. Mark
Artists would find it much easier to paint or draw people	3
Advertisements for hair spray etc. - out	5
No need for Articles for grooming hair:- brushes, comb, rollers, curlers hair-grips, hair-dryers, Brylcreem hair-cream, hair-spray, shampoo	0
hair restorer hair conditioner	2
special hair scissors	3
No Barbers/Hairdressers	0
Confusion	
Would all look same and not be recognized	1
Couldn't tell old from young, boys from girls	3
Police wouldn't identify you	5

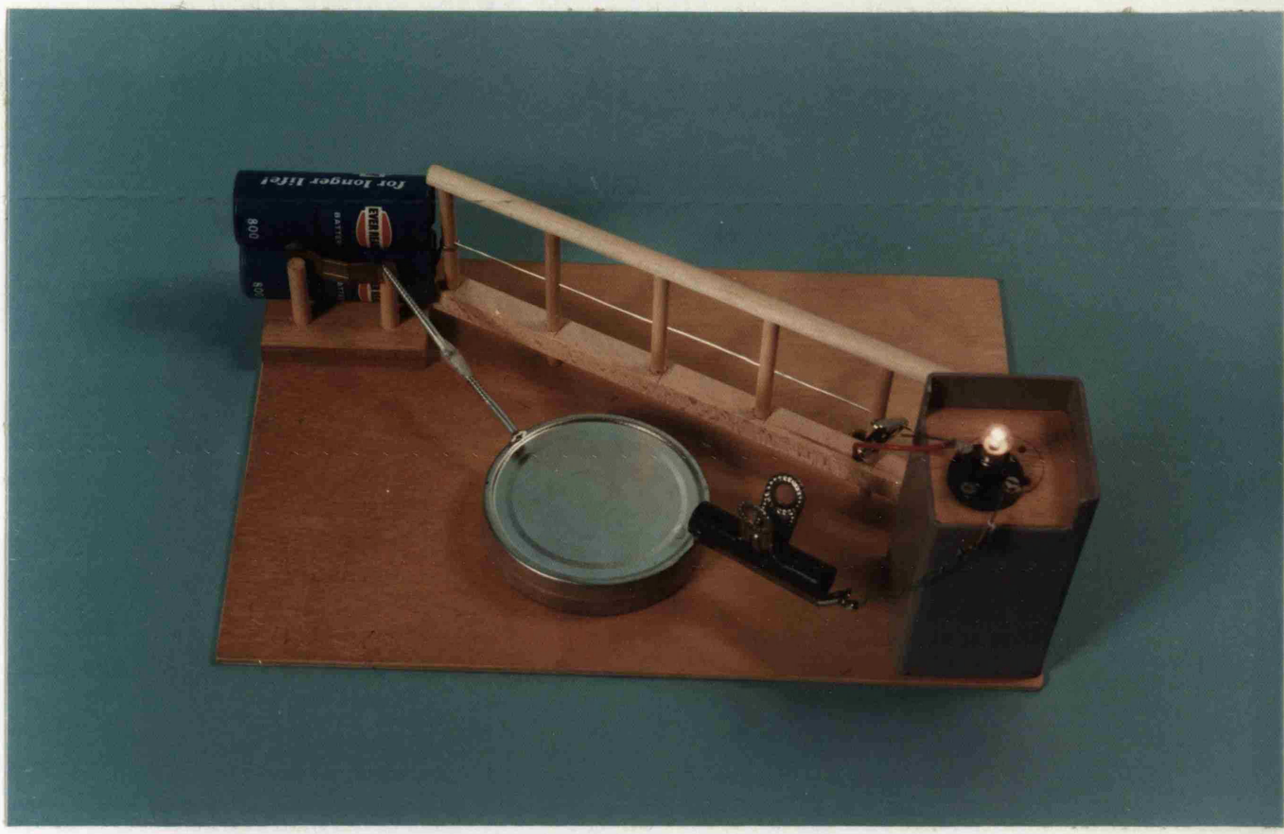
Response	Orig. Mark
Head would get Cold	0
Cool in Summer wouldn't get so hot and sweaty	1
Clown would not need a bald wig	5
Dandruff - none	2
Wouldn't have Dry or greasy hair	3
Eyelashes/Eyebrows gone too	1
Easier to wash head than hair, have to wash our head, easier to keep head clean, difficult to keep head clean	1
Easier to perform operations to head and brain	5
Easier to scratch head	5
Wouldn't get Fleas/Lice	1
Feel ashamed (not go out)	2
Factories would have to turn to making wigs/hats (mass produce)	3

Response	Orig. Mark
Would Glue dogs' hair on, wig on, grass on	2
Graphs of hair colour out	5
Wouldn't go Grey	3
Wouldn't need to Groom hair in some way:- keep hair tidy, put it up, cut it, perm it, put curlers in, dye it, wash it, brush/comb it, have hair-styles (pony-tails, fuzzy hair, curls, plaits)	0
Hair - could not chew it, pull fingers through it	2
Hair would not:- blow in your face get in the way get in your eyes get over your ears be a tangle/have knots	2
Hats needed to keep head warm	2
Have to wear Hats	0
Hats/Wigs sold out	1
Hats/Wigs on easier and off more easily	3
Hats would be too big	5
No need for Hat/ Umbrellas to keep hair tidy/dry	3
No need for swimming Hats	3
Horse Hair would be used more	5
Indians would have no scalp to take	5
Kojak, Kung fu, Yul Brynner would be like the rest of us	0
Kojak wouldn't be so popular, pleased, very happy, would go out of business	1

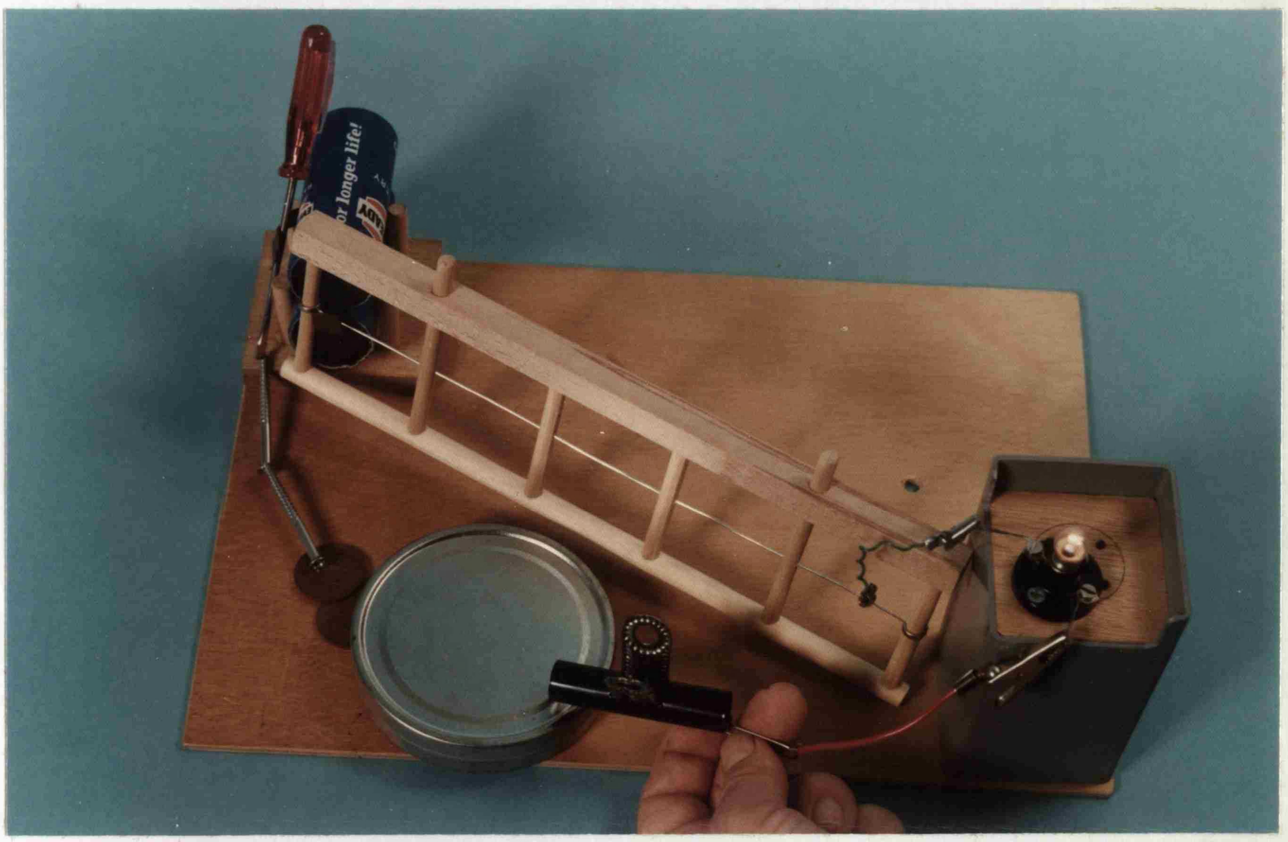
Response	Orig. Mark
We would Look ugly, silly	0
No one would admire Look of friend's hair	5
Wouldn't have trouble of saying "Your hair Looks nice"	5
Laugh at people who Look funny	0
Lollypops would be in demand	2
Save Money/time on haircuts	2
No problems about long/short hair	5
No longhaired hooligans, no hippies	3
No worry about hair, going out to a party, out in the wind	2
Protection for head lost, from injury (would get bruised or cut more easily)	1
Protection for eyes lost (with no eyebrows)	1
Could Polish our heads (instead of brushing)	5
Couldn't Play at hair-dressers	5
R.A.F. wouldn't be known as the 'Brylcream Boys'	5
Razors/shavers, after-shave lotion, shaving cream - all not needed	1
No need to Shave	1
We would Swim/run faster	5
We could Stay in bed longer (no need to do hair)	5
Head would be Sunburnt	3
We would get Sunstroke	5
Shops/firms would lose money/go out of business, lose jobs	0
We would all be Skinheads	3

Response	Orig. Mark
No Torture (children etc. could not pull hair)	1
Teasing people about/ without hair would not be possible	2
Toys would not have real hair	3
Hair would become Valuable	5
It would be Uncomfortable to stand on your head	5
Everyone would buy/ wear Wigs	0
Wig/hat shops would be sold out (see 'Hat')	1
Wig firms would make a lot of money	3
No need for Wigs	3
Wigs would go up in price	3
Couldn't have Wigs, as there would be no hair	5
We would be able to Write shopping lists on our heads	5
We could Wear head lotion to make our heads smell nice	5

Examples of each Category of Response to the Board Game

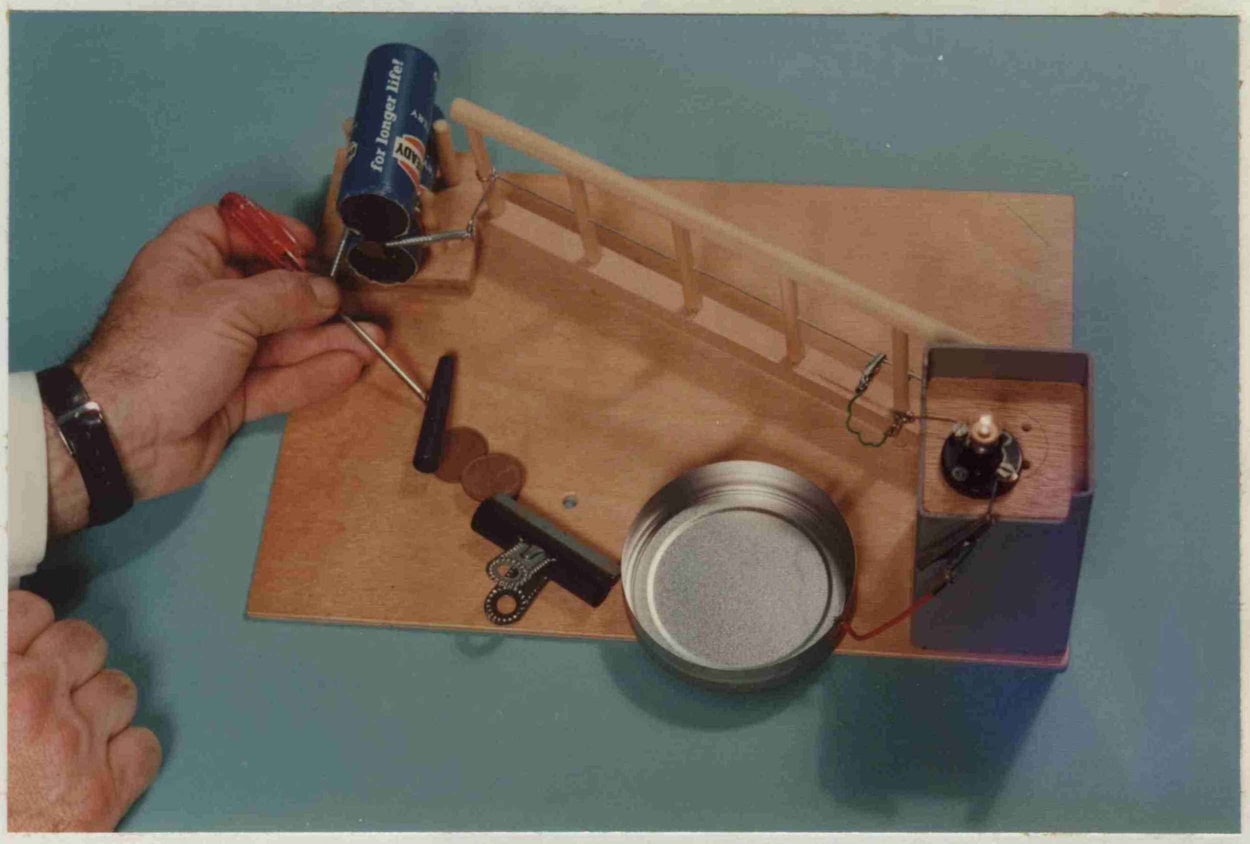


α - response

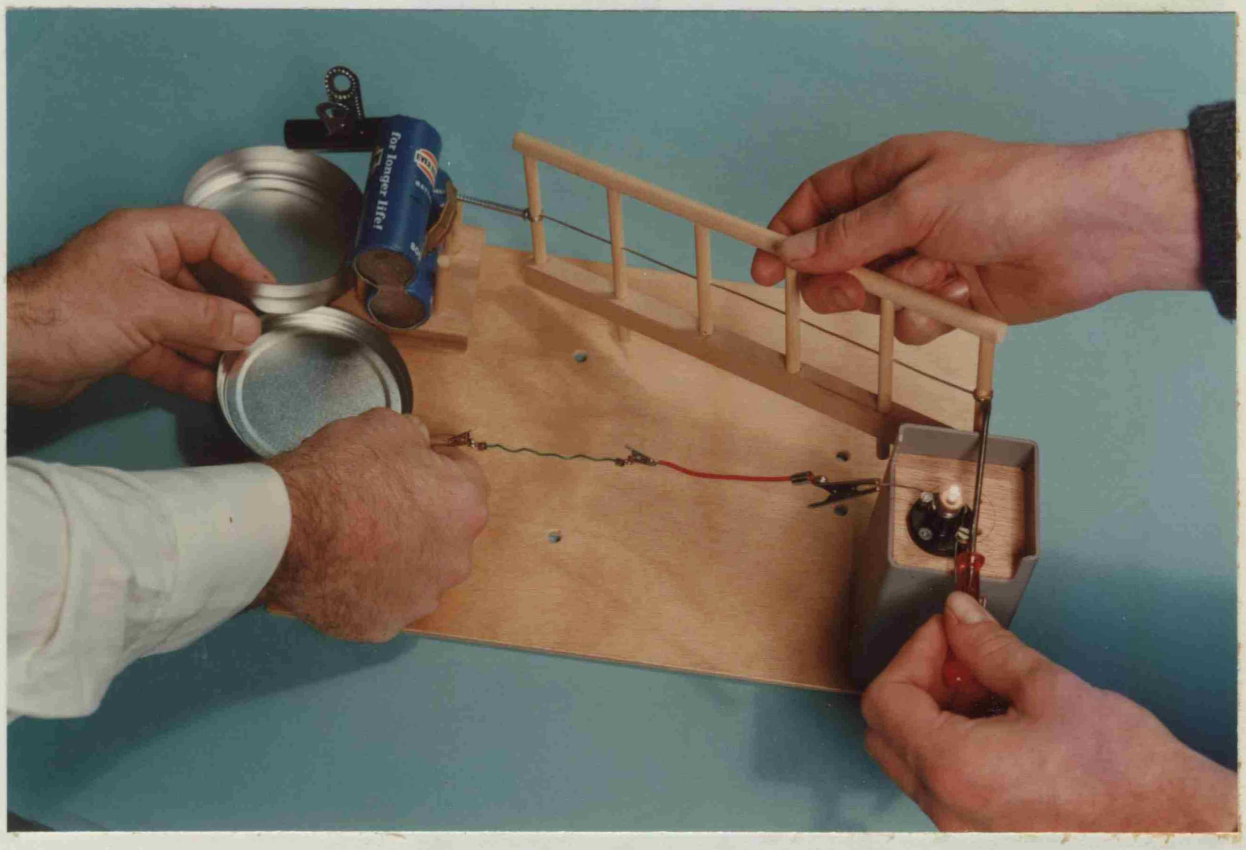


β_1 - response

Responses to the Board Game (cont.)



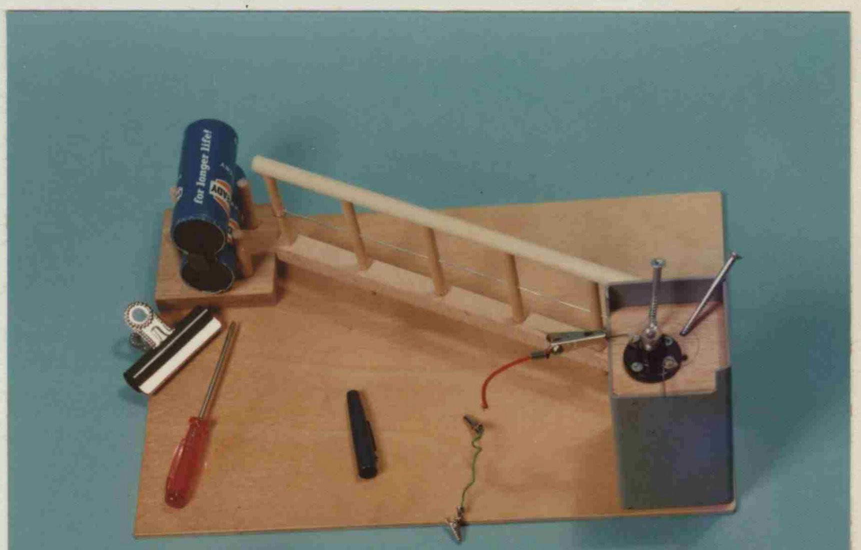
β_2 - response



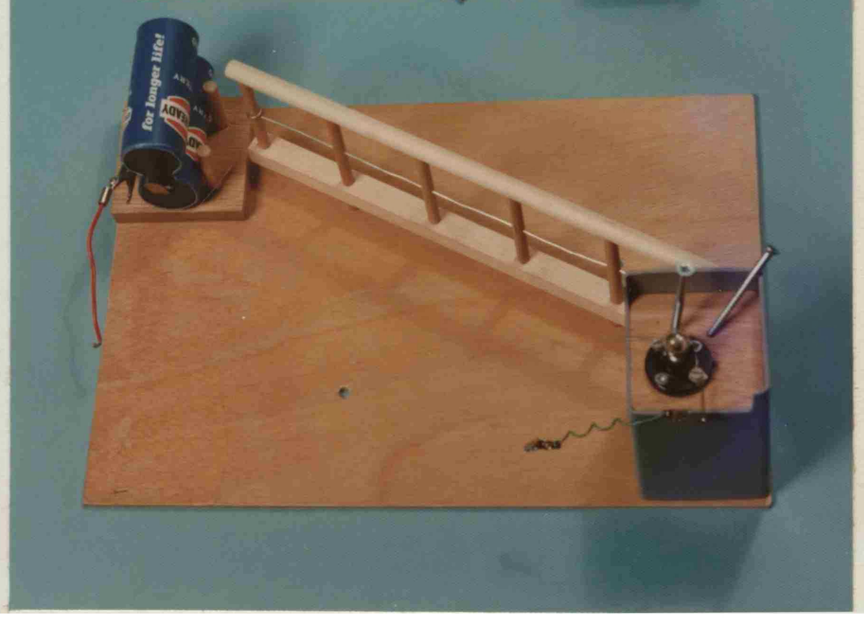
β/γ - response

Responses to the Board Game (cont.)

j_1 - response



j_2 - responses



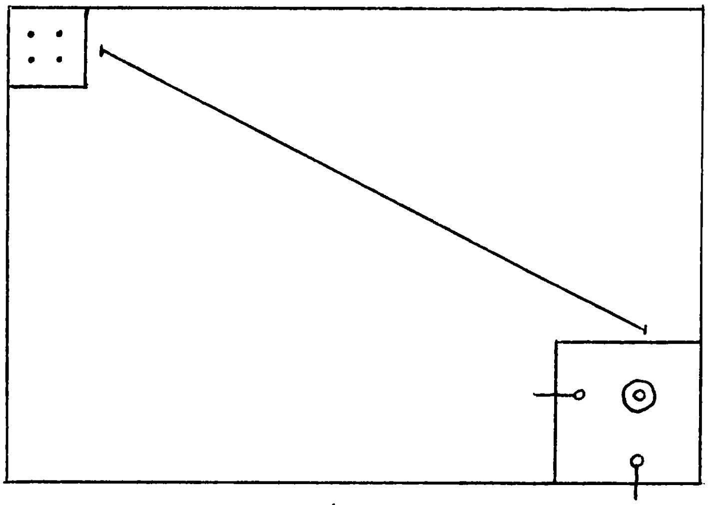
Pupil's 'Report' on the Preliminary Lesson

Electricty

Conductors	Non — conductors
plastic covered wire.	plastic tube
hax saw blade	rubber
perspex	cardboard
metal clamp.	Plywood
copper wire	Foam ball
screw	Polystrene
nail	metal plate
	stone
	button
	Formica
	string
	straw
	wooden rod.
	wool

Today we found out that electricty can pass through some things and not through others. We had a battery, two crocodile clips and a light. We also had 20 other things. All metals will conduct electricty, provided they are clean and not painted. If the electricty is powerfull enough it will go through other things even water.

Name ________________



	Used	Modi-fied?
paper clip		
coil (2 crocs.)		
wire (1 croc.)		
tin		
screws		
pencil		
ruler		
polystyrene		
wood		
sellotape		
straw		
bulldog		
screwdriver		
coins		
pen		
fence		

TIME	Comments
Start	

Other Observations:

1. Changed battery round?
2. Prepared to change ends?
3. Prepared to change round to improve solution?
   at beginning?
   near end?
4. Worked from both ends?
5. (Straight line/small steps/logical) v. (Flair/insight)

Summary

NAME .... SCHOOL ....

This is not an examination, it is part of a SURVEY to find out how good pupils are at thinking up new and interesting ideas.

Some years ago a similar survey was carried out in junior schools and most of you were involved then, now we want to see how you do when you are older.

There are no right or wrong answers so write down as many ideas as you can think of. Work quickly, each part will be timed.

If you need more space continue your answers opposite on the back of the previous page.

DO NOT TURN OVER UNTIL YOU ARE TOLD.

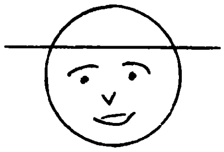
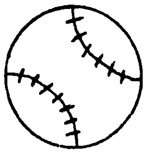
Circles Game

I want to see how many objects you can make from the circles in fifteen minutes. With a pencil add lines to the circles to complete your picture. Your lines can be inside the circle, outside the circle, or both inside and outside.

Make as many DIFFERENT things as you can.

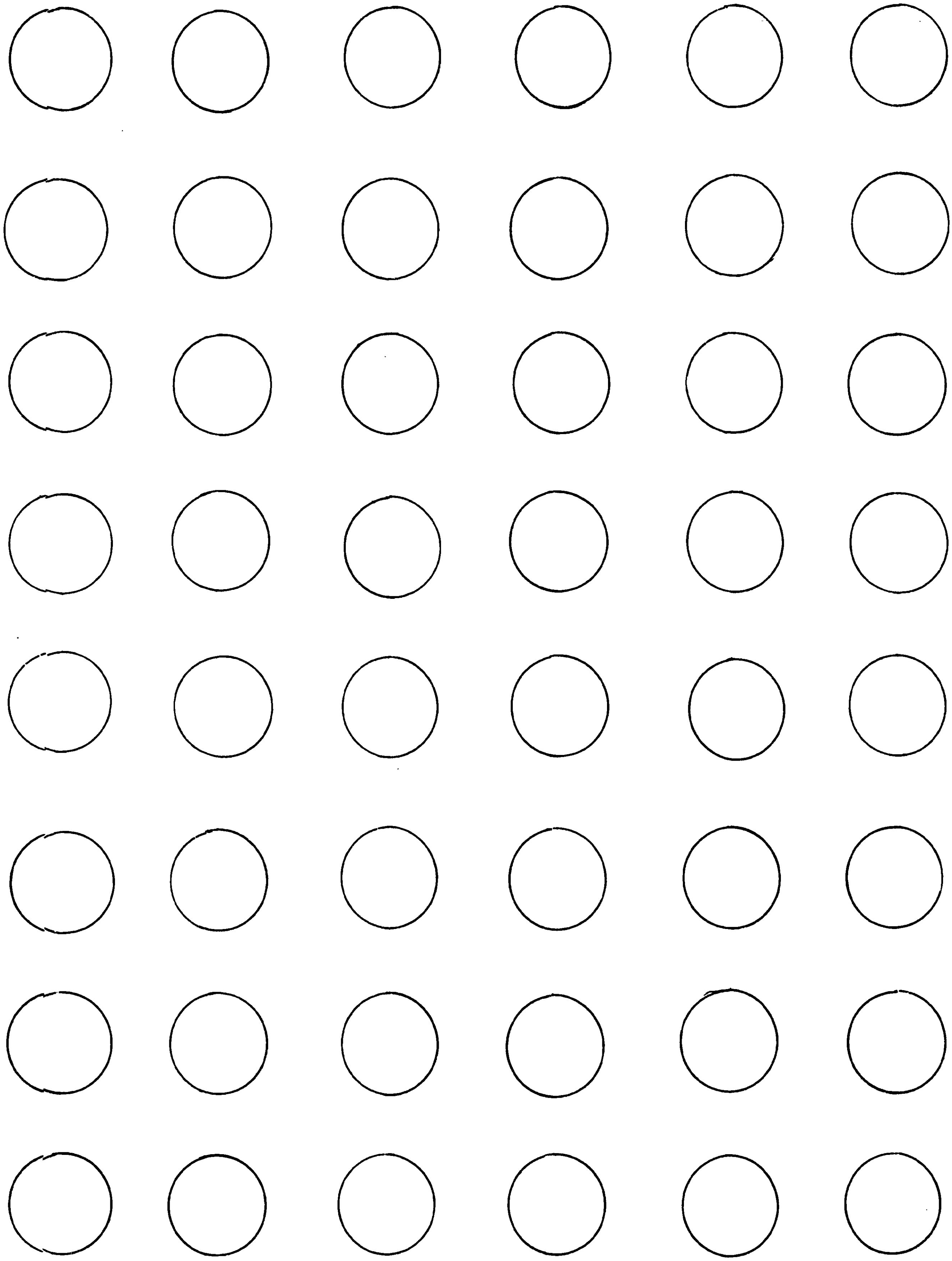
Do not spend much time on any one drawing - you may add titles under your drawings if you do not think they are clear enough.

Look at the two examples on the next page and make as many of your own as you can.

Here are the two examples  

tennis ball

You have 15 minutes to make as many of your own as you can.



If you finish this page draw some more circles opposite.

Uses for Things Time: 15 minutes

The names of THREE objects are written below. I want you to write down as many DIFFERENT uses as you can for each object. Write down anything that comes into your mind, no matter how strange it may seem. Here is an example:-

A BUCKET hold water, sit on, make a helmet with

1 A NEWSPAPER

2 A SPOON

3 A PIECE OF STRING

Consequences Time: 10 minutes

Herre is an example of some things that would be different if everyone had only one hand:-

we could not use a bow and arrow, we might count in fives instead of tens, we would not need a pair of gloves, could not thread a needle.

I wwant you to pretend that the two changes given below suddenly happened. Write down as marny different results of the changes as you can think of.

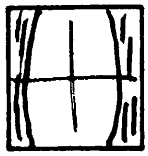
1 If we had no hair on our heads __________

2 If we did not need to eat or drink __________

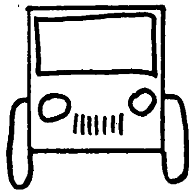
SQUARES GAME

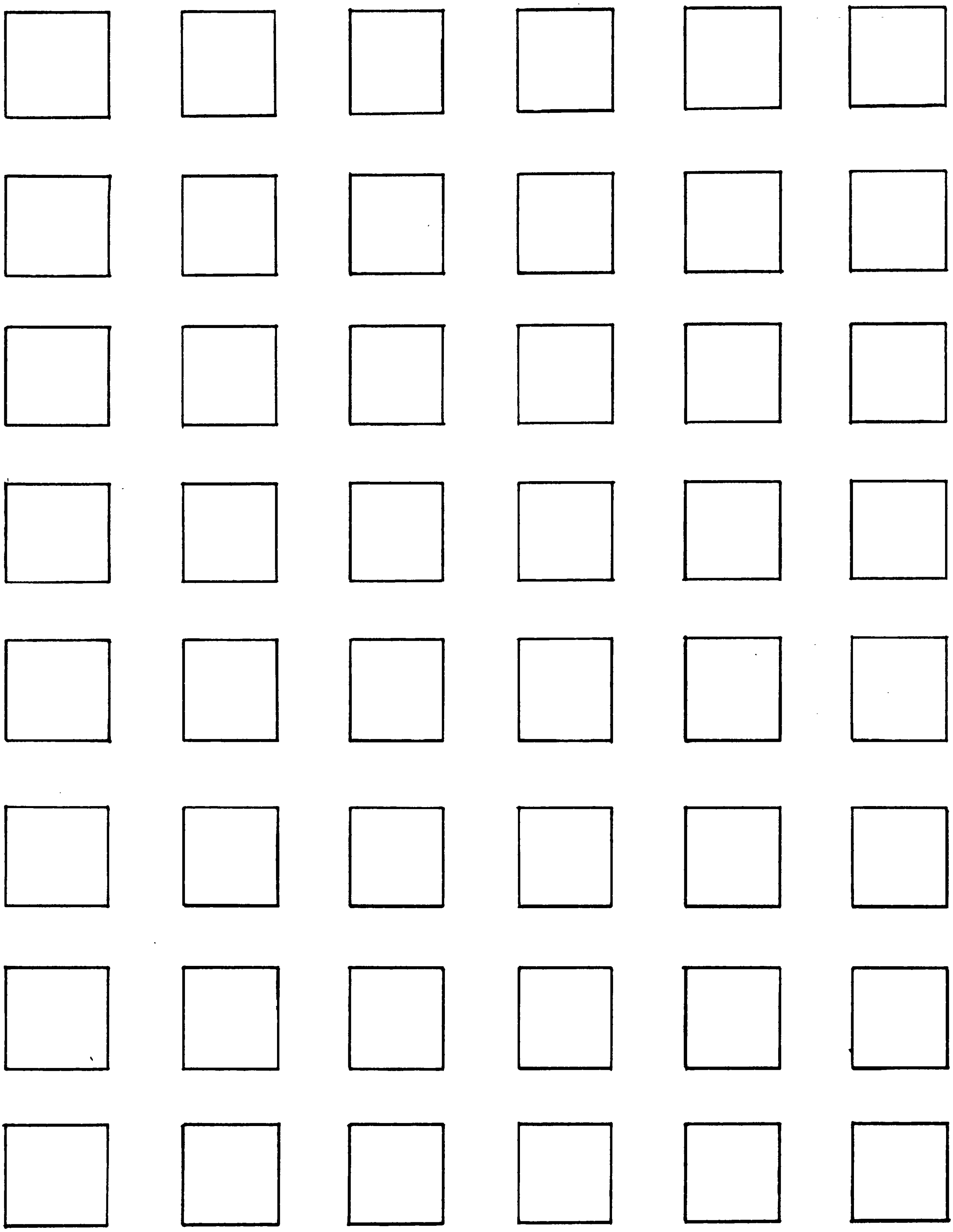
Make as many DIFFERENT objects as you can from the squares below.

Here are two examples:



window





If yyou finish this page ask for another.

WHAT I THINK ABOUT SCHOOL

I have been talking to lots of boys and girls and they have told me things they like and dislike about school.

Over the page you will see some of the things they have said. I should like you to indicate what you feel and think about these things - whether you agree or disagree with what other boys and girls have said. Only I will see your answers.

This is not a test and is not going to be looked at by your school.

There are NO RIGHT and NO WRONG answers.

Just say what you think is most true of you. Please answer as truthfully as you can.

Here is an example:

	YES, OFTEN	SOMETIMES	NEVER
A. I like watching television			

If you often like watching television, put an X in the box marked often.

If you sometimes like watching television, put an X in the box marked sometimes.

Here is one for you to try:

	YES	NOT SURE	NO
B. I like staying in bed late			

		Yes	Not sure	No	
1.	If I missed a games lesson I should be disappointed				
		Always	Sometimes	Never	
2.	I'm sorry when school is over for the day				
		Often	Sometimes	Never	
3.	It's nice to fool about in class			1	
		Most of the time	Sometimes	Hardly ever	
4.	Teacher gets on well with me	1	$\frac{1}{2}$	$-\frac{1}{2}$	
		Yes, often	Sometimes	Hardly ever	
5.	I get a lot of sums wrong		1	2	
		Always	Sometimes	Never	
6.	When the teacher goes out of the room I play about			$\frac{1}{2}$	
		Yes	Not sure	No	
7.	I think I'm pretty good at school work	2	1		
		Most of the time	Sometimes	Never	
8.	School lessons are boring		1	1	
		Yes	Not sure	No	
9.	My class is nicest of all				
		True often	Sometimes	Never	
10.	I have no one to play with at playtime				
		Yes	Not sure	No	
11.	I should like to be better at games than at school work		1	2	
		Yes, often	Sometimes	Hardly ever	
12.	We spend too much time doing arithmetic			$\frac{1}{2}$	
		Yes	Not sure	No	
13.	I'd rather be in my class than the other(s) for my age	2	1		

		Yes, true	Not sure	False		
14.	I sometimes think I'm no good at anything		1	2		
		Yes	Not sure	No		
15.	Other classes think we're nice in my class					
		Yes	Not sure	No		
16.	I think a lot of children of my age would like to be in my class					
		Yes	Not sure	No		
17.	My teacher thinks I'm clever	2	1			
		Yes	Not sure	No		
18.	I bet going out to work is better than school		1	1		
		Yes	Not sure	No		
19.	I shall be sorry to leave my class					
		Yes, often	Sometimes	Never		
20.	I'm scared to ask my teacher for help when I don't understand		1	1		
		Yes, true	Not sure	False		
21.	I have no friends I like very much in my class					
		Yes	Not sure	No		
22.	I like people who get me into mischief			1		
		Yes, often	Sometimes	Never		
23.	I like doing hard sums	1/2				
		Yes	Not sure	No		
4.	Teacher is always nagging me			1		
		Always	Sometimes	Hardly ever		
5.	School is boring			1		
		Yes	Not sure	No		
6.	I'm happy to be in the class I'm in now					

27.	School work worries me	Yes	Not sure	No
			✓	✓
28.	I feel scared when teacher asks me questions about my work	Yes, often	Sometimes	Never
			✓	✓
29.	Other children think we're very clever in my class	Yes, true	Not sure	False
30.	When we have tests I get very good marks	Most of the time	Sometimes	Hardly ever
		2	1	
31.	We have interesting lessons in school	Most of the time	Sometimes	Hardly ever
		1		
32.	Children who can't do their schoolwork feel ashamed	Yes	Not sure	No
				1
33.	I dislike children who are noisy in class	Yes	Not sure	No
		$\frac{1}{2}$		
34.	I hate being in the class I'm in now	Yes	Not sure	No
35.	I like children who get into trouble	Yes	Not sure	No
			1	2
36.	Teacher is interested in me	Yes	Not sure	No
		1		
37.	My class gets blamed for things we don't do	Yes, true	Not sure	False
38.	I should feel a little afraid if I got my spellings or sums wrong	Yes	Not sure	No
			✓	✓
39.	I think the other children in my class like me	Yes	Not sure	No

0.	I'd prefer to be in another class	Yes	Not sure	No
1.	School is fun	Always	Sometimes	Hardly ever
		1		
2.	I find a lot of school work is difficult to understand	Yes, often	Sometimes	Hardly ever
			1	2
3.	I should like to be one of the cleverest pupils in the class	Yes	Not sure	No
		2	1	
4.	I work and try very hard in school	Always	Most of the time	Sometimes
		2	1	
5.	I'm very good at sums	Always	Sometimes	Hardly ever
		2	1	
6.	I don't always get on well with some of the children in my class	Yes, true	Not sure	False
7.	I enjoy most school work	Yes	Not sure	No
		2	1	
8.	Going to school is a waste of time	Yes	Not sure	No
			1	1
9.	My teacher is nice to me	Most of the time	Sometimes	Hardly ever
		1		
0.	I'm useless at school work	Yes, often	Sometimes	Never
			1	2
1.	Teacher thinks I'm a trouble-maker	Yes	Not sure	No
		$-\frac{1}{2}$	$\frac{1}{2}$	1
2.	I should like to be very good at school work	Yes	Not sure	No
		2	1	

53.	Other children make fun of my class	Often	Sometimes	Never		
54.	I think my teacher likes me	Yes	Not sure	No		
		1				
55.	When people ask what class I'm in I always feel happy to tell them	Yes	Not sure	No		
56.	I like school	Yes	Not sure	No		
		1				
57.	I don't seem to be able to do anything really well in school	Yes, true	Not sure	No		
			1	1		
58.	It would bother me if I got my work wrong.	Yes	Not sure	No		
				1		
59.	I like being in my class	Yes	Not sure	No		
60.	I would leave school tomorrow if I could	Yes	Not sure	No		
			1	1		
61.	I enjoy being asked questions by my teacher	Yes	Not sure	No		
		✓	✓			
62.	Other classes think they're better than us	Yes	Not sure	No		
63.	Doing well at school is most important to me	Yes	Not sure	No		
		2	1			
64.	At school they make you do things you don't want to do	Yes	Not sure	No		
				1		

Name: ______________________

A	B	C	D	E	F	G	H	I	J

WHAT I AM LIKE

The words on this page describe what children are like. Read the words next to each number. Put a tick (✓) in ONE space on each line to show whether you think you are that way MOST OF THE TIME or ABOUT HALF THE TIME or HARDLY EVER. There are no right or wrong answers. Just try to say honestly what you are like.

I THINK I AM:

		Most of the Time	About Half the Time	Hardly Ever
1.	neat	3		
2.	a big help at home	3		
3.	bright in school	3		
4.	shy	1		
5.	a pest	1		
6.	very good in art	3		
7.	scared to take chances	1		
8.	full of fun	3		
9.	a hard worker	3		
10.	polite	3		
11.	trying my best	3		
12.	nice looking	3		
13.	lazy	1		
14.	full of questions about new things	3		
15.	going to do well	3		
16.	sad	1		
17.	good in sports	3		
18.	careless	1		
19.	honest	3		
20.	nervous	1		
21.	good at making things	3		
22.	bad	1		
23.	liked by other children	3		
24.	as lucky as others	3		

TEACHER'S RATINGS

CHILDREN'S SCHOOL WORK AND PERSONALITY

Please rate the children in your class on the two aspects of school work and the three personality questions given below.

A. Attainment in school work

1. English

This should be an overall rating of the pupil's attainment in English (including written work, comprehension etc.)

2. Mathematics

This is a rating of the pupil's overall attainment in Mathematics topics (including number work, measurement, shapes etc.)

For these ratings please use the following grades

A	B	C	D	E
very good	good	average	weak	very weak

B. Personality Questions

3. Shy?

4. Lively?

5. Is a pleasure to have in class?

For these questions please use the following grades

A	B	C	D	E
very often	quite often	sometimes	seldom	not at all

Name	Attainment		Personality		
	Eng	Maths	Shy?	Lively?	Is a pleasure in class?

INDICATORS OF CREATIVE BEHAVIOUR

RATING SCALE

Please score the children on each of the characteristics given below. Do not spend too much time assessing the children but enter your impressionistic mark from your experience of teaching the child this year.
The children are listed overleaf, with one column given to each characteristic.
Please enter

2 if the child has shown the characteristic to a great extent
1 " " " " " " " sometimes
0 " " " " " " " very seldom or not at all

CHARACTERISTICS

1. FULL OF IDEAS, ready with suggestions etc.
2. CONSTRUCTS, good at building, making things
3. PERSISTENT, persevering, unwilling to give up
4. IMAGINATIVE, fantasy creating, story telling etc.
5. CURIOUSITY, inquiring, inquisitive, penetrating questions etc.
6. NON-CONFORMING, not bothered by acceptance, or otherwise, by others
7. FLEXIBLE, in ideas and thoughts, not a 1-track mind, tries differing approaches
8. ORIGINAL, unusual answers or solutions, unusual approach to problem solving, artistic creation etc.
9. INDEPENDENT, self-sufficient, individualistic
10. EXPERIMENTER, investigates, tries out new ideas, new equipment etc.

Name	Characteristics										
	1	2	3	4	5	6	7	8	9	10	

Name	Characteristics										
	1	2	3	4	5	6	7	8	9	10	

Name: ....

THINGS I WOULD LIKE TO DO MOST

For each pair of activities given below choose the one you would prefer. Put a tick (✓) in the space next to your choice.

Please do every one. There are no right or wrong answers. So just pick the one you would like to do most.

HERE ARE TWO EXAMPLES:

1. a. Go swimming on a hot day ( )
   b. Dig a ditch on a hot day ( )

2. a. Go swimming on a cold day ( )
   b. Sit by a fire on a cold day ( )

Now do the rest. Just tick the one you would like most.

1. a. Work on a mask ( ) a
   b. Work on a jigsaw puzzle ( )

2. a. Draw a picture ( ) a
   b. Colour in a colouring book ( )

3. a. Write a short story ( ) a
   b. Do a crossword puzzle ( )

4. a. Be a good lawyer ( ) b
   b. Be a good author ( )

5. a. Play cards ( ) b
   b. Act in a play ( )

6. a. Arrange a display in a book shop window ( ) a
   b. Put books on the right shelf in a book shop ( )

7. a. Memorise a poem ( ) b
   b. Write a poem ( )

8. a. Write a short story ( ) a
   b. Practise handwriting ( )

9. a. Memorise the names of the new African countries ( ) b
   b. Make up new names for new countries ( )

10. a. Trace a drawing with tracing paper ( ) b
    b. Paint something on paper ( )

11. a. Be a rich, successful artist ( ) a
    b. Be a rich, successful banker ( )

12. a. Learn to cook a good recipe ( ) b
    b. Think up a new recipe ( )

13. a. Collect stamps ( ) b
    b. Make a model aeroplane ( )

14. a. Play a musical instrument ( ) a
    b. Own a record collection ( )

15. a. Work on a topic chosen by the teacher ( ) b
    b. Work on a topic you choose yourself ( )

16. a. Try to write a good book ( ) a
    b. Try to read a good book ( )

17 a. Write a neat careful account of a visit ( ) b
   b. Write an original story ( )

18. a. Look after the goods in a shop ( ) b
    b. Sell things in a shop ( )

19. a. Make a rock collection ( ) b
    b. Carve a rock into a figure ( )

20. a. Read famous books ( ) a
    b. Learn the names of famous books ( )

21. a. Repeat an important experiment ( ) b
    b. Try a new experiment ( )

22. a. Have the teacher tell you how to do something ( ) b
    b. Figure out how to do something yourself ( )

23. a. Play music ( ) a
    b. Listen to music ( )

24. a. Solve a problem with a group of classmates ( ) b
    b. Try to solve a problem yourself ( )

25. a. Practise spelling ( ) b
    b. Write a story ( )

26. a. Do something you're used to ( ) b
    b. Do something new ( )

27. a. Work with clay ( ) a
    b. Collect post cards ( )

28. a. Be a good shop proprietor ( ) b
    b. Be a good musical composer ( )

29 a. Summarise what is in a book ( ) b
   b. Criticise what is in a book ( )

30. a. Try to invent something ( ) a
    b. Try to fix something ( )

31. a. Learn about something from the textbook ( ) b
    b. Learn about something on your own ( )

32. a. Teach a friend how to play a game ( ) b
    b. Make up a game of your own ( )

33. a. Make a model from a kit ( ) b
    b. Make a model without a kit ( )

34. a. Think up a new way to do something ( ) a
    b. Learn the best way to do something ( )

35. a. Have a job where the boss tells you how to do the work ( ) b
    b. Have a job where you decide how to do some things ( )

36. a. Write a play with some classmates ( ) b
    b. Write a play yourself ( )

Name: ....

THINGS YOU HAVE DONE IN YOUR SPARE TIME

Below is a list of activities boys and girls sometimes do in their spare time. Indicate the ones you have done by putting a tick (✓) in the space at the left. Tick only the things you have done for yourself, NOT the things you have been given or told to do.

( ) 1. Written a poem

( ) 2. Written a story

( ) 3. Written a play

( ) 4. Kept a collection of my writings

( ) 5. Written a song or jingle

( ) 6. Made a puppet show

( ) 7. Kept a diary for at least a month

( ) 8. Played word games with other boys and girls

( ) 9. Used an encyclopedia or some other book in addition to a dictionary

( ) 10. Found mistakes in books or newspapers

( ) 11. Acted in a play or sketch

( ) 12. Made up or organised a play or sketch

( ) 13. Made up and sung a song

( ) 14. Made up a musical composition for some instrument

( ) 15. Made up a new game and taught it to someone else

( ) 16. Acted out a story with others

( ) 17. Written a letter to a member of family or a friend away from home

( ) 18. Made up an original dance

( ) 19. Played at guessing mimes

( ) 20. Visited a zoo

Things done in your spare time (continued)

( ) 21. Explored a cave

( ) 22. Read a science magazine

( ) 23. Read a science book

( ) 24. Mixed colours

( ) 25. Grown crystals

( ) 26. Made a leaf collection

( ) 27. Made a wildflower collection

( ) 28. Made an electric circuit

( ) 29. Made an electric motor

( ) 30. Made a musical instrument

( ) 31. Planned an experiment

( ) 32. Dissected an animal

( ) 33. Grown plants from seeds

( ) 34. Grown plants by taking cuttings

( ) 35. Distilled water

( ) 36. Used a magnifying glass

( ) 37. Made ink

( ) 38. Made leaf prints

( ) 39. Started a fire with a lens

( ) 40. Used a magnet

( ) 41. Raised rats, mice, rabbits, or guinea pigs

( ) 42. Collected insects

( ) 43. Collected rocks

( ) 44. Kept a daily record of weather

( ) 45. Been a bird watcher

( ) 46. Kept a science notebook

( ) 47. Kept a science scrapbook

( ) 48. Attended a science fair or display

( ) 49. Used a chemistry set

( ) 50. Produced static electricity

Things done in your spare time (continued)

( ) 51. Constructed a model aeroplane

( ) 52. Counted annual rings in a log

( ) 53. Made a stamp collection

( ) 54. Made a collection of post marks

( ) 55. Organised or helped to organise a club

( ) 56. Served as officer in a club organised by boys and/or girls

( ) 57. Figured out a way of improving a game we play at school or home

( ) 58. Figured out a way of improving the way we do something at home

( ) 59. Figured out a way of improving the way we do something at school

( ) 60. Figured out a way of improving the way we do something in a club, Scouts, etc.

( ) 61. Solved a problem about getting along with other boys and girls

( ) 62. Solved a problem about getting along with my parents

( ) 63. Helped act out some historical event

( ) 64. Found out about the history of my city or community

( ) 65. Found out about the way some government agency (post office, court, etc.) operates

( ) 66. Wrote a letter to someone in another country

( ) 67. Made a map of my street or district

( ) 68. Made my own decision about the use of money

( ) 69. Asked questions about the way some business works

( ) 70. Made a poster for some club, school or other event

( ) 71. Organised or helped organise paper drive, jumble sale, etc.

( ) 72. Sketched landscape with pencil and/or charcoal

( ) 73. Designed stage settings for play or sketch

( ) 74. Made up a design for jewellry

( ) 75. Made up a design for cloth

( ) 76. Illustrated a story of my own or one in a book

( ) 77. Taken colour photographs

Things done in your spare time (continued)

( ) 78. Taken black and white photographs

( ) 79. Made plaster objects from a mould

( ) 80. Drawn cartoons

( ) 81. Designed greeting card for Christmas or special event.

( ) 82. Made linoleum cuts

( ) 83. Made block prints in colour

( ) 84. Made a watercolour painting of a familiar scene

( ) 85. Made an oilcolour painting of some type

( ) 86. Made things from papier maché

( ) 87. Made a toy for a child

( ) 88. Built a scale model of a park, playground, farm, etc.

( ) 89. Made a wood carving

( ) 90. Made a soap carving

( ) 91. Made a basket for ornamental purposes

( ) 92. Drawn up plans for an invention, apparatus etc.

( ) 93. Constructed a model of an invention

( ) 94. Made up recipe for some kind of food dish (meat, salad, dessert, etc.)

( ) 95. Made up recipe for some kind of drink mixture.

INTERESTS.

A.

Below you will see some of the things boys and girls of your age find interesting.
You are to give marks to the things *you* like doing either in or out of school.

Here is an example : Swimming ☐

If you like swimming *very much*, put a 2 in the box.
If you *quite* like swimming, put a 1 in the box.
If you *don't like* swimming or have *never tried* it, put a 0 in the box.

Like very much = 2 Quite like = 1 Never tried it or don't like it = 0

Now give marks to the activities below making sure you put a mark in each box.

1.	Watching TV	
2.	Writing stories	I
3.	Playing football	
4.	Drawing	G
5.	Reading comics	
6.	Making up plays ..	I
7.	Reading adventure stories	
8.	Going for walks	
9.	Collecting stamps ..	G
10.	Making up poems ..	I
11.	Playing cricket	
12.	Sewing	G
13.	Reading the newspaper ..	
14.	Painting pictures ..	I
15.	Climbing trees	

(Please turn over and continue)

16.	Acting	I
17.	Playing marbles	
18.	Playing the recorder or other musical instrument	I
19.	Going to the cinema ..	
20.	Gardening	G
21.	Reading poetry	I
22.	Dancing	G
23.	Playing chess	L
24.	Doing science experiments	L
25.	Playing hopscotch ..	
26.	Making models	L
27.	Doing crossword puzzles	G
28.	Writing a daily diary ..	G
29.	Playing hide and seek ..	
30.	Reading encyclopedias ..	L

B.

1. When you are not at school what hobby do you like *best* of all ?

....

2. What do you think you would like to do when you grow up ?

....

....